

**UNIVERSIDAD AUTONOMA DE MADRID**

**ESCUELA POLITECNICA SUPERIOR**



## **TRABAJO FIN DE GRADO**

### **APROXIMACIONES UNIFORMES PARA DISTRIBUCIONES DE PROBABILIDAD DE SUMAS DE VARIABLES ALEATORIAS POSITIVAS CON COLAS PESADAS**

**Ana Cristina Navarro Rodríguez**  
**Tutor: Alberto Suárez González**

**Julio 2014**



## Resumen

En nuestro entorno encontramos constantemente fenómenos que pueden ser descritos utilizando modelos probabilísticos. Este conocimiento sobre las características de la evolución del sistema nos permite predecir futuras situaciones y progresar en las distintas disciplinas donde tienen lugar estos eventos modelizados.

El inconveniente se presenta cuando, a pesar de conocer la distribución que rige el suceso, sus funciones de densidad de probabilidad y de distribución acumulada no tienen una forma explícita cerrada que sea conocida. Surge de aquí la necesidad de encontrar aproximaciones con fórmula analítica cerrada a sumas de variables aleatorias positivas con colas pesadas, objetivo del presente trabajo.

Cuando el número de variables de la suma ( $N$ ) tiende a infinito, el teorema del límite central nos indica que la distribución de la suma de variables aleatorias independientes con varianza finita se aproxima a una Gaussiana. Por ello, cuando  $N$  es suficientemente grande podemos utilizar esta distribución para modelizar la suma de variables aleatorias. Sin embargo, nos encontramos frente a un problema cuando  $N$  no cumple esta propiedad.

Para solucionar este problema y encontrar una aproximación con alta precisión se proponen las series de Gram-Charlier y Edgeworth, basadas en los momentos de la distribución. Inicialmente se estudiará el modelo clásico de estas series, se ampliará con su correspondiente generalización y finalmente se trabajará sobre modificaciones en las expresiones de ambas series.

Para valorar la calidad de las estimaciones propuestas se realizarán experimentos sobre sumas de variables lognormales. Aparte de su interés práctico en campos como las finanzas y las telecomunicaciones, la ventaja de analizar este tipo de variables aleatorias es que, si bien presentan colas pesadas, todos sus momentos son finitos. En la realización de estas pruebas se empleará código desarrollado en MATLAB, debido a su adecuación para realizar análisis matemático.

Como conclusión se extrae que las series generalizadas y la modificación propuesta aproximan satisfactoriamente a sumas de variables cuando las colas no son excesivamente pesadas o cuando  $N$  es grande. Además, en todos los casos, las aproximaciones dadas por la truncación de la serie de Edgeworth son mejores que las correspondientes a la serie de Gram Charlier.

**Palabras clave:** aproximación, sumas de variables aleatorias, función de densidad de distribución, serie de Gram-Charlier, serie de Edgeworth.

## Abstract

In the surrounding environment they are constantly found phenomena that can be described using probabilistic models. This knowledge about the system evolution characteristics enable to predict future situations and evolve in different disciplines where this modelled events take place.

The problem arises when, despite knowing the distribution governing the event, its probability density functions and cumulative distribution does not present a known explicit closed form. Thus, the need arises to find approaches with analytical closed formula to sums of positive heavy-tailed random variables, being the aim of this work.

When the number of variables of the sum ( $N$ ) tends to infinity, the central limit theorem indicates that the distribution of the sum of independent random variables with finite variance approaches a Gaussian distribution. Therefore, when  $N$  is large enough, such a distribution can be used to model the sum of random variables. However, an issue is encountered when  $N$  does not satisfy this property.

In order to solve this problem and find out a high precision approximation, Gram-Charlier and Edgeworth expansions, based on the moments of the distribution, are proposed. Initially, the classical model of this series will be studied. This will be extended with the corresponding generalization and finally changes in the expressions of both series will be analysed.

To assess the quality of the proposed estimations, experiments will be performed on sums of log-normal variables. In addition to its application interest in fields as finance and telecommunications, the advantage of analysing this kind of random variables is that, even they present heavy tails, all their moments are finite. For the accomplishment of these trials, a code developed in MATLAB will be used due to its suitability for performing mathematical analysis.

As a conclusion, it is drawn that the generalized series and the proposed modification successfully approximate to sums of variables when the tails are not excessively heavy or when  $N$  is big enough. Furthermore, in every case, the approximations given by the truncation of Edgeworth expansion exhibit an improvement over the results obtained by the Gram-Charlier expansion.

**Keywords:** approach, sum of random variables, probability density function, Gram-Charlier expansion, Edgeworth expansion.

# Índice

<b>Lista de figuras</b> . . . . .	<b>V</b>
<b>Abreviaturas</b> . . . . .	<b>VII</b>
<b>1 Introducción</b> . . . . .	<b>1</b>
1.1 Motivación . . . . .	1
1.2 Objetivos . . . . .	2
1.3 Estructura del documento . . . . .	2
<b>2 Estado del arte</b> . . . . .	<b>5</b>
2.1 Aproximación mediante una distribución conocida . . . . .	5
2.2 Series de aproximación . . . . .	6
<b>3 Sumas de variables aleatorias</b> . . . . .	<b>7</b>
3.1 Teorema del límite central . . . . .	7
3.2 Serie de Gram Charlier . . . . .	9
3.2.1 Gram Charlier desarrollo con propiedad de ortogonalidad . . . . .	9
3.2.2 Gram-Charlier desarrollo alternativo . . . . .	12
3.2.3 Generalización de la serie de Gram-Charlier . . . . .	15
3.3 Serie de Edgeworth . . . . .	17
3.3.1 Edgeworth . . . . .	17
3.3.2 Generalización de la serie de Edgeworth . . . . .	21
3.4 Modificación de las series de Gram-Charlier y Edgeworth . . . . .	24
3.4.1 Modificación de la serie de Gram-Charlier . . . . .	24
3.4.2 Modificación de la serie de Edgeworth . . . . .	25
3.5 Transformada de Fourier para pdf de sumas de variables aleatorias . . . . .	27
<b>4 Librería desarrollada</b> . . . . .	<b>29</b>
4.1 Documentación de las funciones de la librería . . . . .	29
<b>5 Experimentos numéricos</b> . . . . .	<b>33</b>
5.1 Versión 1 . . . . .	35
5.1.1 Suma de 10 variables . . . . .	35
5.1.2 Suma de 50 variables . . . . .	39
5.2 Versión 2 . . . . .	42
5.2.1 Suma de 10 variables . . . . .	43
5.2.2 Suma de 50 variables . . . . .	46
5.3 Versión 3 . . . . .	49
5.3.1 Suma de 10 variables . . . . .	49
5.3.2 Suma de 50 variables . . . . .	53
<b>6 Conclusiones</b> . . . . .	<b>57</b>
6.1 Conclusiones y contribuciones . . . . .	57
6.2 Trabajo futuro . . . . .	58
<b>Anexos</b> . . . . .	<b>61</b>

<b>A</b>	<b>Cumulantes</b>	<b>61</b>
<b>B</b>	<b>Polinomios de Hermite</b>	<b>64</b>
<b>C</b>	<b>Delta de Dirac</b>	<b>66</b>
	<b>Referencias</b>	<b>67</b>

## Índice de figuras

1	Red de celdas. . . . .	1
2	Ejemplo histograma y pdf calculada con transformadas de Fourier rápidas .	28
3	Versión 1 suma de 10 $\text{LogN}(0, 0.3)$ . . . . .	36
4	Versión 1 suma de 10 $\text{LogN}(0, 0.5)$ . . . . .	37
5	Versión 1 suma de 10 $\text{LogN}(0, 1)$ . . . . .	38
6	Versión 1 suma de 50 $\text{LogN}(0, 0.3)$ . . . . .	39
7	Versión 1 suma de 50 $\text{LogN}(0, 0.5)$ . . . . .	40
8	Versión 1 suma de 50 $\text{LogN}(0, 1)$ . . . . .	41
9	Versión 2 suma de 10 $\text{LogN}(0, 0.3)$ . . . . .	43
10	Versión 2 suma de 10 $\text{LogN}(0, 0.5)$ . . . . .	44
11	Versión 2 suma de 10 $\text{LogN}(0, 1)$ . . . . .	45
12	Versión 2 suma de 50 $\text{LogN}(0, 0.3)$ . . . . .	46
13	Versión 2 suma de 50 $\text{LogN}(0, 0.5)$ . . . . .	47
14	Versión 2 suma de 50 $\text{LogN}(0, 1)$ . . . . .	48
15	Versión 2 suma de 10 $\text{LogN}(0, 0.3)$ . . . . .	50
16	Versión 2 suma de 10 $\text{LogN}(0, 0.5)$ . . . . .	51
17	Versión 2 suma de 10 $\text{LogN}(0, 1)$ . . . . .	52
18	Versión 2 suma de 50 $\text{LogN}(0, 0.3)$ . . . . .	53
19	Versión 2 suma de 50 $\text{LogN}(0, 0.5)$ . . . . .	54
20	Versión 2 suma de 50 $\text{LogN}(0, 1)$ . . . . .	55
21	Error relativo absoluto de la versión 2 suma de 10 $\text{LogN}(0, 0.5)$ . . . . .	58





## Abreviaturas

<b>pdf</b>	probability density function
<b>cdf</b>	cumulative distribution function
<b>ccdf</b>	cumulative distribution function
<b>GC</b>	Gram-Charlier
<b>Edg</b>	Edgeworth



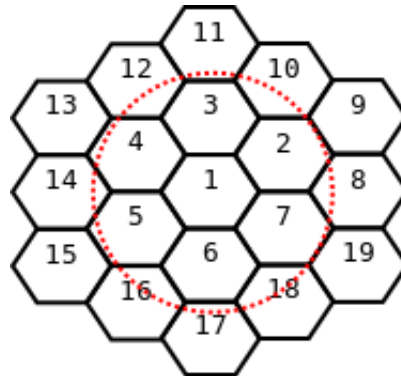
# 1. Introducción

## 1.1. Motivación

En múltiples áreas el ser humano intenta modelizar fenómenos que le permitan predecir o anticipar futuros comportamientos, así como hacer posible un proceso de mejora y avance continuo en dispareas disciplinas. De ahí el interés por el análisis de las distribuciones de sumas de variables aleatorias positivas, independientes y con colas pesadas, objetivo de las investigaciones descritas en este documento. Dos ejemplos que permiten apreciar el amplio universo de aplicaciones prácticas de modelización de sumas de este tipo de variables son: los campos de la comunicación móvil y las finanzas.

La comunicación móvil es un área de gran relevancia en la actualidad, la cual permite la comunicación entre dos terminales sin cable. Una red de comunicaciones móviles puede modelizarse de manera esquemática como una red de celdas. En cada una de las celdas (denominadas estaciones base) se encuentra una antena repetidora (transmisor) encargada de propagar la señal recibida a otros transmisores hasta llegar al terminal receptor.

En estas redes la transmisión puede ser fallida, provocando por tanto una caída del servicio (outage), y es para determinar la probabilidad de este fallo para lo que se necesitan sumas de variables lognormales independientes [11]. Se trata de una suma ya que se ven involucradas todas las celdas incluidas en un radio determinado.



**Figura 1:** Red de celdas.

Una ciencia social de interés son las finanzas, encargada de cubrir el análisis y valoración de productos derivados. Entre la amplia tipología de los mismos se encuentran las opciones, contratos entre dos partes que dan derecho (no obligación) a comprar/vender un determinado bien a un cierto precio (strike) en/hasta un cierto tiempo (vencimiento). Aquellas que dan derecho a compra reciben el nombre de call y las que dan derecho a venta son denominadas put.

Un tipo conocido de opciones son las asiáticas con media aritmética. Éstas se caracterizan porque el cálculo de su precio toma en consideración precisamente la media aritmética de los valores del activo en el mercado en diferentes momentos a lo largo de la vida de la

opción. Los precios del activo son modelizados mediante una lognormal (modelo Black-Scholes), es por ello que en la distribución de la media vuelve a jugar un papel fundamental la suma de variables lognormales independientes [4] y [8].

El caso particular de la lognormal es ampliamente utilizado en áreas tan dispares como las mencionadas. Su suma no tiene forma explícita y por ello es objeto de estudio de muchos investigadores [9] y [10].

## 1.2. Objetivos

Este trabajo fin de grado tiene como objetivo final aproximar mediante series distribuciones de probabilidad de sumas de variables aleatorias positivas con colas pesadas. Más detalladamente, para conseguir dicho propósito se cuenta con los siguientes objetivos parciales:

- Estudio de las series clásicas de Gram-Charlier y Edgeworth actualmente empleadas como estimadores de funciones de densidad de probabilidad.
- Examinar la generalización de las expansiones de Gram-Charlier y Edgeworth. Estas series presentan la posibilidad de mejorar las series clásicas adecuándose a la distribución que se desee aproximar.
- Desarrollo de una librería en MATLAB con la implementación de las distintas series analizadas.
- Realización de pruebas sobre casos particulares elaborando un estudio sobre la evolución y ajuste de las distintas aproximaciones.

## 1.3. Estructura del documento

Este documento está dividido en seis secciones. En la segunda sección a modo de iniciación al objetivo principal de este trabajo se presentan diversos métodos utilizados para la aproximación de sumas de variables.

En la tercera sección se encuentra el estudio realizado sobre las sumas de variables aleatorias, incluyendo el teorema del límite central, el desarrollo exhaustivo de las diferentes series de aproximación de Gram-Charlier y Edgeworth, y por último el uso de la transformada de Fourier para estimar numéricamente la pdf de una suma de variables aleatorias.

En la cuarta sección se describe la librería desarrollada con las series de Gram-Charlier y Edgeworth.

La quinta sección está dedicada a la experimentación sobre las series estudiadas en la sección 3. En ella se observan y comentan los resultados obtenidos con cada una de ellas.

Finalmente, en la última sección se indican las conclusiones extraídas, los aportes realizados y posible trabajo futuro.



## 2. Estado del arte

La aproximación a distribuciones de sumas de v.a. positivas independientes con colas pesadas ha sido un campo muy estudiado desde su aparición hasta hoy en día, momento en el que se sigue intentando encontrar un método para aproximar de la manera más ajustada y con el menor coste computacional posible.

### 2.1. Aproximación mediante una distribución conocida

Una de las ideas sobre como aproximar una suma de v.a.i.i.d. es intentar ajustar una familia paramétrica de distribuciones cuya forma funcional es conocida. Para ello se seleccionan los valores de los parámetros de la distribución que maximicen alguna medida de bondad de ajuste.

Cuando se trata de una suma de lognormales, la distribución que se elige para aproximar puede ser una lognormal, y se realizan distintos tipos de ajuste en los parámetros de la media y la varianza de la nueva distribución. En el libro "Principles of mobile communication", de Gordon L. Stüber [11] podemos encontrar tres métodos diferentes que se utilizan para encontrar la distribución de ésta única variable aproximadora.

1. Método de Fenton-Wilkinson: consiste en hacer coincidir los dos primeros momentos de la nueva variable con los correspondientes a la suma, esto es, elige los dos parámetros de los que consta una distribución lognormal de tal forma que tengan misma media y varianza.
2. Método de Schwartz y Yeh: consiste en hacer coincidir los dos primeros momentos del logaritmo de la variable aproximadora con los del logaritmo de la suma de variables. Se realiza de forma recursiva, tomando en primer lugar las dos primeras variables de la suma, y una vez que encuentra su correspondiente aproximación repite el proceso tomando la aproximación hallada y la tercera variable de la suma. Esta estrategia se efectúa hasta finalizar con todos los sumandos, obteniendo entonces una nueva variable lognormal que aproxima a la suma completa.
3. Método de Farley: este método difiere en los anteriores al no consistir en hacer coincidir los primeros momentos de dos variables. El método de Farley lo que realiza es una aproximación a la función de distribución acumulada complementaria de la suma de variables, que se define como 1 menos la función de distribución acumulada.

En el artículo [9] establecen otro método de asignación de parámetros. En su caso igualan los dos primeros momentos de la inversa de la suma con los de la inversa de la distribución lognormal. Comparando con los métodos de Fenton-Wilkinson y Schwartz y Yeh encuentran una aproximación más precisa.

También se realizan otros tipos de *matching* para aproximar la cola de la distribución, haciendo coincidir pares de momentos distintos al primero y segundo, tomando por ejem-

plo el segundo y tercer momento o incluso el tercero y cuarto. Esto lo encontramos en el artículo "Generalized Gram-Charlier series with application to the sum of log-normal variates", de D.C. Scheleher [10].

## 2.2. Series de aproximación

Otra idea de aproximación a las sumas de v.a.i.i.d. es aproximar su función de densidad mediante series. Hallamos entonces las expansiones de Gram-Charlier y de Edgeworth, las cuales típicamente tienen como función de densidad de referencia la gaussiana. Las series se expresan en función de los momentos de la suma de variables.

La expansión de Edgeworth es una reordenación de los términos de la serie de Gram-Charlier, con la que normalmente se obtienen mejores resultados, ya que en este reordenamiento se controla el orden de magnitud de los términos, mientras que en Gram-Charlier el orden de los términos no decrece monótonamente [1]. Sin embargo en ambas series, la mejor aproximación al truncar la serie se suele obtener tomando un número pequeño de términos.

En [10] podemos observar que se realiza una generalización de la serie de Gram-Charlier para que tenga cualquier función de densidad como función de referencia. En su caso, toman aproximaciones basadas en igualar pares de momentos, como indicamos anteriormente. En próximas subsecciones veremos que la serie de Edgeworth también se puede generalizar para poder cambiar la función de referencia y se estudiará el desarrollo de ambas.



### 3. Sumas de variables aleatorias

#### 3.1. Teorema del límite central

El teorema central del límite es ampliamente conocido y utilizado al tratar con sumas de variables aleatorias para aproximarlas con la distribución normal.

Antes de mostrar el enunciado del teorema y su demostración, donde nos apoyaremos en el artículo [3], veamos la definición de convergencia en distribución.

##### Definición

Sea una sucesión  $\{X_n\}$  y  $X$  v.a. con cdf  $\{F_n\}$  y  $F$  respectivamente, se dice que  $\{X_n\}$  converge en distribución a  $X$ , y se denota como  $X_n \xrightarrow{d} X$ , si  $\forall x$  donde  $F$  es continua se tiene:

$$\lim_{n \rightarrow \infty} F_n(x) = F(x). \quad (1)$$

Veamos pues el enunciado del teorema y su demostración apoyándonos en el artículo [3], donde estudian la expansión de Edgeworth y convergencia en el teorema central del límite.

##### Teorema

Sea  $X_n$  una sucesión de v.a.i.i.d. con media  $\mu$  y varianza  $\sigma^2$  distinta de cero, y sea

$$Y_n = \sum_{j=1}^n X_j.$$

Entonces

$$\frac{Y_n - E(Y_n)}{\sqrt{V(Y_n)}} \xrightarrow{d} N(0, 1), \quad (2)$$

donde  $E$  es el operador esperanza y  $V$  el operador varianza.

##### Demostración

En primer lugar definimos  $S_n = \frac{Y_n - E(Y_n)}{\sqrt{V(Y_n)}} = \frac{Y_n - n\mu}{\sigma\sqrt{n}}$ .

Sea  $\hat{\phi}_n(t)$  la función característica de  $S_n$ :

$$\hat{\phi}_n(t) = E(e^{itS_n}) = E\left(\prod_{j=1}^n e^{it\frac{(X_j - \mu)}{\sigma\sqrt{n}}}\right) = E\left(e^{it\frac{(X_j - \mu)}{\sigma\sqrt{n}}}\right)^n = \hat{\phi}\left(\frac{t}{\sigma\sqrt{n}}\right)^n,$$

$\hat{\phi}$  tiene segunda derivada continua y utilizando el desarrollo de Taylor de la exponencial  $\left(e^x = \sum_{m=0}^{\infty} x^m/m!\right)$  cuando  $t \rightarrow 0$  tenemos:

$$\hat{\phi}(t) = E\left(e^{it(X_1-\mu)}\right) = \hat{\phi}(0) + t\hat{\phi}'(0) + \frac{t^2}{2}\hat{\phi}''(0) + o(t^2) = 1 - \frac{t^2\sigma^2}{2} + o(t^2), \quad (3)$$

ya que

$$\begin{aligned} \hat{\phi}(0) &= 1, \\ \hat{\phi}'(0) &= i(E(X) - \mu) = 0, \\ \hat{\phi}''(0) &= -E((X - \mu)^2) = -\sigma^2. \end{aligned} \quad (4)$$

Si fijamos  $t$ , empleando (3)

$$\hat{\phi}\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{\sigma^2 n}\right), \text{ cuando } n \rightarrow \infty. \quad (5)$$

Por lo tanto tenemos que la función característica de  $S_n$

$$\hat{\phi}_n(t) = \hat{\phi}\left(\frac{t}{\sigma\sqrt{n}}\right)^n \longrightarrow e^{-\frac{t^2}{2}}, \text{ cuando } n \rightarrow \infty. \quad (6)$$

Dado que una distribución  $N(0, 1)$  tiene función característica  $e^{-\frac{t^2}{2}}$  podemos utilizar el Teorema de continuidad de Lévy-Cramer [2], obteniendo así el resultado deseado,

$$S_n = \frac{Y_n - E(Y_n)}{\sqrt{V(Y_n)}} \xrightarrow{d} N(0, 1)$$

■

Si tenemos una suma de v.a.i.i.d. el teorema central del límite nos asegura que la distribución  $N(0,1)$  es muy buena aproximación cuando el número de variables de la suma tiende a infinito, sea cual sea la distribución de las variables siempre y cuando su varianza sea finita. Es por ello que, para un número de sumandos suficientemente grande se utiliza la distribución normal como una estimación.

El problema se tiene cuando no hay suficientes variables, estudiaremos por este motivo series asintóticas como aproximadores, las cuales no necesitan necesariamente que el número de sumandos sea muy elevado para obtener una estimación ajustada.

### 3.2. Serie de Gram Charlier

Para la obtención de la serie de Gram-Charlier veremos dos desarrollos distintos con mismos resultados. En ambos, denotaremos como  $Y$  a la variable aleatoria para la cual deseamos estimar su función de densidad de probabilidad  $f_Y(y)$ .

#### 3.2.1. Gram Charlier desarrollo con propiedad de ortogonalidad

Las pautas generales para este desarrollo las podemos encontrar en [12], aquí veremos más detalladamente los pasos a seguir.

Supongamos en este primer análisis, que  $Y$  tiene media 0 y varianza 1. Siendo  $\phi(y) = \frac{1}{\sqrt{2\pi}}e^{-y^2/2}$  la función de densidad de probabilidad gaussiana, expresamos  $f_Y(y)$  como una perturbación de  $\phi(y)$ :

$$f_Y(y) = \sum_{n=0}^{\infty} c_n (-1)^n \frac{d^n}{dy^n} \phi(y) \quad (7)$$

Podemos ahora utilizar la definición de los polinomios de Hermite (75),

$$H_n(y) = \frac{1}{\phi(y)} (-1)^n \frac{d^n}{dy^n} \phi(y), \quad (8)$$

en la ecuación (7),

$$f_Y(y) = \sum_{n=0}^{\infty} c_n H_n(y) \phi(y). \quad (9)$$

Para calcular los coeficientes  $c_n$  de la serie emplearemos la propiedad de ortogonalidad de los polinomios de Hermite (77), que es enunciada el anexo B.

Multiplicamos ambos lados de la ecuación (9) por  $H_m(y)$ , integramos respecto a  $y$  y utilizamos (77),

$$\begin{aligned} \int_{-\infty}^{\infty} H_m(y) f_Y(y) dy &= \int_{-\infty}^{\infty} H_m(y) \left( \sum_{n=0}^{\infty} c_n H_n(y) \phi(y) \right) dy \\ &= \sum_{n=0}^{\infty} c_n \left( \int_{-\infty}^{\infty} H_m(y) H_n(y) \phi(y) dy \right) = \sum_{n=0}^{\infty} c_n n! \delta_{mn} = c_m m! \end{aligned}$$

$$\Rightarrow c_m = \frac{1}{m!} \int_{-\infty}^{\infty} H_m(y) f_Y(y) dy, \quad m \in \mathbb{R} \quad (10)$$

Utilizando este resultado en la ecuación (9),  $f_Y(y)$  queda expresada de la siguiente manera:

$$f_Y(y) = \sum_{n=0}^{\infty} H_n(y) \phi(y) \frac{1}{n!} \int_{-\infty}^{\infty} H_n(y) f_Y(y) dy. \quad (11)$$

Los cuatro primeros coeficientes  $c_n$  de la serie son:

$$c_0 = \frac{1}{0!} \int_{-\infty}^{\infty} H_0(y) f_Y(y) dy = \int_{-\infty}^{\infty} f_Y(y) dy = 1,$$

$$c_1 = \frac{1}{1!} \int_{-\infty}^{\infty} H_1(y) f_Y(y) dy = \int_{-\infty}^{\infty} y f_Y(y) dy = \mu = 0,$$

$$c_2 = \frac{1}{2!} \int_{-\infty}^{\infty} (y^2 - 1) f_Y(y) dy = \frac{1}{2}(\mu_2 - 1) = 0,$$

$$c_3 = \frac{1}{3!} \int_{-\infty}^{\infty} (y^3 - 3y) f_Y(y) dy = \frac{1}{3!}(\mu_3 - 3\mu) = \frac{\mu_3}{3!},$$

$$c_4 = \frac{1}{4!} \int_{-\infty}^{\infty} (y^4 - 6y^2 + 3) f_Y(y) dy = \frac{1}{4!}(\mu_4 - 6\mu_2 + 3) = \frac{1}{4!}(\mu_4 - 3),$$

donde  $\mu_n$  denota el momento no central de orden  $n$  ( $E[X^n]$ ).

Obtenemos como resultado la pdf de  $X$  en función de sus momentos, los polinomios de Hermite y la  $\phi$  gaussiana:

$$f_Y(y) = \phi(y) \left( 1 + \frac{1}{3!} \mu_3 H_3(y) + \frac{1}{4!} (\mu_4 - 3) H_4(y) + \dots \right) \quad (12)$$

Podemos también expresarla en función de los cumulantes de  $Y$ , cuya definición y propiedades son presentadas en el apéndice A. En esta ocasión,  $k_1 = 0$  y  $k_2 = 1$  dado que hemos tomado una variable con media cero y varianza 1. Utilizando las equivalencias entre cumulantes y momentos no centrales se obtiene:

$$f_Y(y) = \phi(y) \left( 1 + \frac{1}{3!} k_3 H_3(y) + \frac{1}{4!} k_4 H_4(y) + \dots \right) \quad (13)$$

Para una variable  $Y$  cualquiera, que no tenga necesariamente media  $\mu = 0$  y desviación típica  $\sigma = 1$ , se puede obtener la serie de Gram-Charlier siguiendo el mismo procedimiento realizado teniendo en cuenta que ahora en (7) en vez de tomar la función  $\phi(y)$ , emplearemos la pdf de una distribución normal con parámetros  $\mu$  y  $\sigma$ ,

$$\phi(y, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}},$$

y que los polinomios de Hermite cumplen lo siguiente:

$$(-1)^n \frac{d^n}{dy^n} \phi(y, \mu, \sigma) = \frac{1}{\sigma^n} H_n\left(\frac{y-\mu}{\sigma}\right) \phi(y, \mu, \sigma), \quad (14)$$

Obtendremos entonces como ecuación análoga a (9),

$$f_Y(y) = \sum_{n=0}^{\infty} \frac{c_n}{\sigma^n} H_n\left(\frac{y-\mu}{\sigma}\right) \phi(y, \mu, \sigma).$$

Continuando con el desarrollo logramos calcular la función de densidad de  $Y$  como la siguiente serie:

$$f_Y(y) = \phi(y, \mu, \sigma) \left( 1 + \frac{1}{3!\sigma^3} k_3 H_3\left(\frac{y-\mu}{\sigma}\right) + \frac{1}{4!\sigma^4} k_4 H_4\left(\frac{y-\mu}{\sigma}\right) + \dots \right) \quad (15)$$

### 3.2.2. Gram-Charlier desarrollo alternativo

Para obtener la serie de Gram-Charlier podemos realizar un desarrollo alternativo, en el cual se emplea la transformada de Fourier y el operador derivada. Este método nos ayudará a desarrollar la serie de Gram-Charlier generalizada y entender el empleado en la serie de Edgeworth.

En este análisis, partiremos de que la variable  $Y$  tendrá media  $E(Y) = \mu$  y varianza  $V(Y) = \sigma^2$ .

#### Proposición

Sea  $Y$  una variable aleatoria, su correspondiente función de densidad  $f_Y(y)$  se puede expresar como:

$$f_Y(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} E[e^{itY}] dt \quad (16)$$

#### Demostración

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} E[e^{itY}] dt &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} \left( \int_{-\infty}^{\infty} e^{ity'} f_Y(y') dx' \right) dt = \\ &= \int_{-\infty}^{\infty} f(y') \left( \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{it(y'-y)} dt \right) dy' = \\ &= \int_{-\infty}^{\infty} f(y') \delta(y' - y) dy' = f_Y(y) \end{aligned}$$

■

Por la definición de cumulantes tenemos  $\log(E(e^{tY})) = \sum_{n=1}^{\infty} \frac{t^n}{n!} k_n$ . Por lo tanto,  $E(e^{itY})$  puede ser expresado como

$$E(e^{itY}) = \exp \left( \sum_{l=1}^{\infty} \frac{(it)^l}{l!} k_l \right) \quad (17)$$

Por otro lado, tenemos que  $e^{-ity}$  es autofunción del operador  $D_y$  con autovalor  $-it$ ,

$$D_y e^{-ity} = -it e^{-ity}. \quad (18)$$

Utilizando (17),  $k_1 = \mu$ ,  $k_2 = \sigma^2$  y (18), se puede desarrollar la ecuación de la pdf (16) como sigue:

$$\begin{aligned}
f_Y(y) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} E(e^{itY}) e^{-ity} dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left( \sum_{l=1}^{\infty} \frac{(it)^l}{l!} k_l \right) e^{-ity} dt = \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left( \sum_{l=3}^{\infty} \frac{k_l}{l!} (it)^l \right) e^{-ity} e^{it\mu - (s\sigma)^2/2} dt = \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left( \sum_{l=3}^{\infty} \frac{k_l (-1)^l D_y^l}{l!} \right) e^{-ity} e^{it\mu - (s\sigma)^2/2} dt = \\
&= \exp \left( \sum_{l=3}^{\infty} \frac{k_l}{l!} (-1)^l D_y^l \right) \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{t(i\mu - iy) - (t\sigma)^2/2} dt = \\
&= \exp \left( \sum_{l=3}^{\infty} \frac{k_l}{l!} (-1)^l D_y^l \right) \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \left( t\sigma - \frac{i(\mu-y)}{\sigma} \right)^2} e^{-\frac{(\mu-y)^2}{2\sigma^2}} dt = \\
&= \exp \left( \sum_{l=3}^{\infty} \frac{k_l}{l!} (-1)^l D_y^l \right) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\mu-y)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2} dt = \\
&= \exp \left( \sum_{l=3}^{\infty} \frac{k_l}{l!} (-1)^l D_y^l \right) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\mu-y)^2}{2\sigma^2}}
\end{aligned}$$

$$f_Y(y) = \exp \left( \sum_{l=3}^{\infty} \frac{k_l}{l!} (-1)^l D_y^l \right) \phi(y, \mu, \sigma) \quad (19)$$

Utilizando el desarrollo de Taylor de la exponencial  $\left( e^x = \sum_{m=0}^{\infty} x^m/m! \right)$  y la propiedad de los polinomios de Hermite (78) sobre la ecuación (19), se obtiene:

$$\begin{aligned}
f_Y(y) &= \left( 1 + \sum_{l=3}^{\infty} \frac{k_l}{l!} (-1)^l D_y^l + \left( \sum_{l=3}^{\infty} \frac{k_l}{l!} (-1)^l D_y^l \right)^2 \frac{1}{2!} + \dots \right) \phi(y, \mu, \sigma) \\
&= \left( \phi(y, \mu, \sigma) + \sum_{l=3}^{\infty} \frac{k_l}{l!} (-1)^l D_y^l \phi(y, \mu, \sigma) + \frac{1}{2!} \left( \sum_{l=3}^{\infty} \frac{k_l}{l!} (-1)^l D_y^l \phi(y, \mu, \sigma) \right)^2 + \dots \right) \\
&= \left( 1 + \sum_{l=3}^{\infty} \frac{k_l}{l! \sigma^l} H_l \left( \frac{y-\mu}{\sigma} \right) + \sum_{l=3}^{\infty} \sum_{m=3}^{\infty} \frac{k_l k_m}{2! l! m! \sigma^{l+m}} H_{l+m} \left( \frac{y-\mu}{\sigma} \right) + \dots \right) \phi(y, \mu, \sigma)
\end{aligned}$$

Finalmente, al tomar los primeros términos, conseguimos la misma serie de Gram-Charlier del apartado anterior (15):

$$f_Y(y) = \phi(y, \mu, \sigma) \left( 1 + \frac{k_3}{3!\sigma^3} H_3\left(\frac{y-\mu}{\sigma}\right) + \frac{k_4}{4!\sigma^4} H_4\left(\frac{y-\mu}{\sigma}\right) + \dots \right), \quad (20)$$

donde debemos recordar que  $\mu$  es la media de  $Y$  y  $\sigma$  su desviación típica.



### 3.2.3. Generalización de la serie de Gram-Charlier

Como hemos visto en los anteriores apartados, podemos realizar una estimación de la pdf de una variable aleatoria mediante una serie de Gram-Charlier, donde se utilizaba la pdf correspondiente a una distribución normal como base de la serie. En este apartado veremos una generalización, para poder utilizar una función de densidad de cualquier otra distribución.

Los pasos clave para esta generalización los podemos encontrar en [10], donde utilizan esta serie para aproximar la cdf de una suma de variables lognormales.

Tomamos  $Z$  una variable aleatoria con pdf  $g_z$ , la que deseemos como base para la aproximación. Denotaremos como  $k_l^{[Z]}$  a los cumulantes de  $Z$  y como  $k_l^{[Y]}$  a los correspondientes a  $Y$ , variable de la que se desea calcular la función de densidad de probabilidad.

Partiremos de la proposición (16),

$$f_Y(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} E(e^{itY}) e^{-ity} dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left( \sum_{l=1}^{\infty} \frac{(it)^l}{l!} k_l^{[Y]} \right) e^{-ity} dt.$$

Introducimos en la ecuación los cumulantes de  $Z$ ,

$$\begin{aligned} f_Y(y) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left( \sum_{l=1}^{\infty} \frac{k_l^{[Y]} - k_l^{[Z]} + k_l^{[Z]}}{l!} (it)^l \right) e^{-ity} dt = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left( \sum_{l=1}^{\infty} \frac{k_l^{[Y]} - k_l^{[Z]}}{l!} (it)^l \right) \exp \left( \sum_{l=1}^{\infty} \frac{k_l^{[Z]}}{l!} (it)^l \right) e^{-ity} dt, \end{aligned}$$

empleando el operador derivada (18) resulta

$$f_Y(y) = \exp \left( \sum_{l=1}^{\infty} \frac{k_l^{[Y]} - k_l^{[Z]}}{l!} (-1)^l D_y^l \right) \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left( \sum_{l=1}^{\infty} \frac{k_l^{[Z]}}{l!} (it)^l \right) e^{-ity} dt$$

y volviendo a utilizar la proposición (16) conseguimos la pdf de  $Y$  en términos de la pdf de  $Z$ , la función de referencia:

$$f_Y(y) = \exp \left( \sum_{l=1}^{\infty} \frac{k_l^{[Y]} - k_l^{[Z]}}{l!} (-1)^l D_y^l \right) g_Z(y)$$

Ahora podemos expresar la exponencial con su desarrollo de Taylor y ordenando según el orden de derivada se tiene la serie de Gram-Charlier generalizada:

$$f_Y(y) = g_Z(y) + \sum_{m=1}^{\infty} \frac{c_m}{m!} (-1)^m D_y^m g_Z(y) \quad (21)$$

donde los cuatro primeros términos de la serie son:

$$\begin{aligned} c_1 &= \left( k_1^{[Y]} - k_1^{[Z]} \right), \\ c_2 &= \left( k_2^{[Y]} - k_2^{[Z]} \right) + \left( k_1^{[Y]} - k_1^{[Z]} \right)^2, \\ c_3 &= \left( k_3^{[Y]} - k_3^{[Z]} \right) + 3 \left( k_1^{[Y]} - k_1^{[Z]} \right) \left( k_2^{[Y]} - k_2^{[Z]} \right) + \left( k_1^{[Y]} - k_1^{[Z]} \right)^3, \\ c_4 &= \left( k_4^{[Y]} - k_4^{[Z]} \right) + 4 \left( k_1^{[Y]} - k_1^{[Z]} \right) \left( k_3^{[Y]} - k_3^{[Z]} \right) + 3 \left( k_2^{[Y]} - k_2^{[Z]} \right)^2 + \\ &\quad + 6 \left( k_1^{[Y]} - k_1^{[Z]} \right)^2 \left( k_2^{[Y]} - k_2^{[Z]} \right) + \left( k_1^{[Y]} - k_1^{[Z]} \right)^4. \end{aligned}$$

Si elegimos la variable aleatoria  $Z$  tal que coincida en media y varianza con  $Y$ , dado que  $k_i = \mu$  y  $k_2 = \sigma^2$  los cuatro primeros coeficientes de (21) serán:

$$c_1 = c_2 = 0, \quad c_3 = \left( k_3^{[Y]} - k_3^{[Z]} \right), \quad c_4 = \left( k_4^{[Y]} - k_4^{[Z]} \right),$$

siendo por tanto,

$$f_Y(y) = g_Z(y) - \frac{\left( k_3^{[Y]} - k_3^{[Z]} \right)}{3!} D_y^3 g_Z(y) + \frac{\left( k_4^{[Y]} - k_4^{[Z]} \right)}{4!} D_y^4 g_Z(y) + \dots \quad (22)$$

### 3.3. Serie de Edgeworth

#### 3.3.1. Edgeworth

La serie de Edgeworth es como la de Gram-Charlier pero con un reordenamiento diferente de los términos. Para obtener la serie de Edgeworth en primer lugar se supone que se quiere aproximar una función de densidad correspondiente a  $S_n$  una suma de v.a., se procede al cálculo de la serie con su correspondiente ordenación en función de  $n$  y finalmente se tomará  $n = 1$ , obteniendo así una aproximación a una pdf de una única variable aleatoria.

Sean  $Y_j$  v.a.i.i.d. entonces  $S_n$  se define como:

$$S_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n Y_j \quad (23)$$

Por la proposición (16) tenemos,

$$f_{s_n}(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} E(e^{itS_n}) dt \quad (24)$$

Donde  $E(e^{itS_n})$  puede ser expresado en función de la esperanza de  $Y = Y_1$ ,

$$E(e^{itS_n}) = E\left(\prod_{j=1}^n e^{it\frac{Y_j}{\sqrt{n}}}\right) = \prod_{j=1}^n E\left(e^{it\frac{Y_j}{\sqrt{n}}}\right) = E\left(e^{it\frac{Y}{\sqrt{n}}}\right)^n \quad (25)$$

Introducimos esta igualdad en (24) y desarrollamos de igual manera que lo hacíamos para

la serie de Gram-Charlier,

$$\begin{aligned}
f_{s_n}(y) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} E \left[ e^{it \frac{y}{\sqrt{n}}} \right]^n dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \exp \left( n \sum_{l=1}^{\infty} \frac{(it)^l}{n^{\frac{l}{2}} l!} k_l^{[Y]} \right) dt \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left( n \sum_{l=3}^{\infty} \frac{(it)^l}{n^{\frac{l}{2}} l!} k_l^{[Y]} \right) e^{-itx} e^{it\mu^{[Y]} \sqrt{n} - t^2 \sigma^{[Y]2}/2} dt \\
&= \exp \left( n \sum_{l=3}^{\infty} \frac{k_l^{[Y]}}{n^{\frac{l}{2}} l!} (-1)^l D_y^l \right) \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{t(i\mu^{[Y]} \sqrt{n} - iy) - t^2 \sigma^{[Y]2}/2} dt \\
&= \exp \left( n \sum_{l=3}^{\infty} \frac{k_l^{[Y]}}{n^{\frac{l}{2}} l!} (-1)^l D_y^l \right) \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2} (t^2 \sigma^{[Y]2} - 2t(i\mu^{[Y]} \sqrt{n} - iy))} dt \\
&= \exp \left( n \sum_{l=3}^{\infty} \frac{k_l^{[Y]}}{n^{\frac{l}{2}} l!} (-1)^l D_y^l \right) \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \left( t\sigma - \frac{(i\mu^{[Y]} \sqrt{n} - iy)}{\sigma} \right)^2 - \frac{(\mu^{[Y]} \sqrt{n} - y)^2}{2\sigma^{[Y]2}}} dt \\
&= \exp \left( n \sum_{l=3}^{\infty} \frac{k_l^{[Y]}}{n^{\frac{l}{2}} l!} (-1)^l D_y^l \right) \frac{1}{\sigma^{[Y]} \sqrt{2\pi}} e^{-\frac{(\mu^{[Y]} \sqrt{n} - y)^2}{2\sigma^{[Y]2}}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt \\
&= \exp \left( n \sum_{l=3}^{\infty} \frac{k_l^{[Y]}}{n^{\frac{l}{2}} l!} (-1)^l D_y^l \right) \frac{1}{\sigma^{[Y]} \sqrt{2\pi}} e^{-\frac{(y - \mu^{[Y]} \sqrt{n})^2}{2\sigma^{[Y]2}}} \\
&= \exp \left( n \sum_{l=3}^{\infty} \frac{k_l^{[Y]}}{n^{\frac{l}{2}} l!} (-1)^l D_y^l \right) \phi(y, \mu^{[Y]} \sqrt{n}, \sigma^{[Y]}) \\
&= \phi(y, \mu^{[Y]} \sqrt{n}, \sigma^{[Y]}) \left[ 1 + \right. \\
&\quad + n \sum_{l=3}^{\infty} \frac{k_l^{[Y]}}{n^{\frac{l}{2}} l! \sigma^{[Y]l}} H_l \left( \frac{y - \mu^{[Y]} \sqrt{n}}{\sigma^{[Y]}} \right) + \\
&\quad \left. + n^2 \sum_{l=3}^{\infty} \sum_{m=3}^{\infty} \frac{k_l^{[Y]} k_m^{[Y]}}{n^{\frac{l+m}{2}} 2! l! m! \sigma^{[Y]l+m}} H_{l+m} \left( \frac{y - \mu^{[Y]} \sqrt{n}}{\sigma^{[Y]}} \right) + \dots \right]
\end{aligned} \tag{26}$$

Ahora desarrollando según las potencias de  $n$ ,

$$\begin{aligned}
f_{S_n}(y) = & \phi(y, \mu^{[Y]} \sqrt{n}, \sigma^{[Y]}) \left[ 1 + \right. \\
& + \frac{1}{\sqrt{n}} \left( \frac{k_3^{[Y]}}{3! \sigma^{[Y]3}} H_3(\hat{y}) \right) + \\
& + \frac{1}{\sqrt{n}^2} \left( \frac{k_4^{[Y]}}{4! \sigma^{[Y]4}} H_4(\hat{y}) + \frac{k_3^{[Y]2}}{2! 3! 3! \sigma^{[Y]6}} H_6(\hat{y}) \right) + \\
& + \frac{1}{\sqrt{n}^3} \left( \frac{k_5^{[Y]}}{5! \sigma^{[Y]5}} H_5(\hat{y}) + \frac{k_3^{[Y]} k_4^{[Y]}}{3! 4! \sigma^{[Y]7}} H_7(\hat{y}) + \frac{k_3^{[Y]3}}{3! 4! \sigma^{[Y]9}} H_9(\hat{y}) \right) + \\
& \left. + \dots \right],
\end{aligned} \tag{27}$$

donde  $\hat{y} = \left( \frac{y - \mu^{[Y]} \sqrt{n}}{\sigma^{[Y]}} \right)$ .

Para finalizar tomamos  $n = 1$ , de esta forma  $S_n = Y_1 = Y$ , entonces los primeros términos de la serie serán:

$$\begin{aligned}
f_Y(y) = & \phi(y, \mu, \sigma) \left[ 1 + \right. \\
& + \left( \frac{k_3}{3! \sigma^3} H_3\left(\frac{y-\mu}{\sigma}\right) \right) + \\
& + \left( \frac{k_4}{4! \sigma^4} H_4\left(\frac{y-\mu}{\sigma}\right) + 10 \frac{k_3^2}{6! \sigma^6} H_6\left(\frac{y-\mu}{\sigma}\right) \right) + \\
& + \left( \frac{k_5}{5! \sigma^5} H_5\left(\frac{y-\mu}{\sigma}\right) + 35 \frac{k_3 k_4}{7! \sigma^7} H_7\left(\frac{y-\mu}{\sigma}\right) + 280 \frac{k_3^3}{9! \sigma^9} H_9\left(\frac{y-\mu}{\sigma}\right) \right) + \\
& \left. + \dots \right],
\end{aligned} \tag{28}$$

donde  $\mu$ ,  $\sigma^2$  y  $k_l$  son la media, varianza y cumulantes de  $Y$ .

Como podemos observar se trata en realidad de una reordenación diferente de los coeficientes de la serie de Gram-Charlier, gracias a este reordenamiento nos encontramos ante

una serie que en algunos casos mejora la aproximación de Gram-Charlier, como se contempla en [5] y como veremos más adelante en algunos resultados prácticos.

Cabe también destacar que las truncaciones de orden 0 y orden 3 de la serie de Edgeworth coinciden con las de Gram-Charlier, comenzando a diferir a partir del cuarto término.

### 3.3.2. Generalización de la serie de Edgeworth

Al igual que generalizamos la expansión de Gram-Charlier, también es posible generalizar la serie de Edgeworth, para poder utilizar diferentes funciones base y trabajar con la más adecuada según la distribución de la que vayamos a calcular la función de densidad de probabilidad.

Comenzaremos como en el apartado anterior, suponiendo

$$S_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n Y_j \quad (29)$$

donde  $Y_j$  son v.a.i.i.d., y denotando a  $Y_1$  como  $Y$ .

Utilizando la proposición (16) y la relación entre la función generadora de momentos de  $S_n$  e  $\frac{Y}{\sqrt{n}}$  (25) partimos de:

$$\begin{aligned} f_{s_n}(y) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} E(e^{itS_n}) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} E \left[ e^{it \frac{Y}{\sqrt{n}}} \right]^n dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \exp \left( n \sum_{l=1}^{\infty} \frac{(it)^l}{n^{\frac{l}{2}} l!} k_l^{[Y]} \right) dt \end{aligned} \quad (30)$$

Tomamos una variable aleatoria  $Z$ , con pdf  $g_Z$  y cumulantes  $k_l^{[Z]}$ , y los introducimos en (31),

$$\begin{aligned} f_{s_n}(y) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left( n \sum_{l=1}^{\infty} \frac{k_l^{[Y]} - k_l^{[Z]} + k_l^{[Z]}}{n^{\frac{l}{2}} l!} (it)^l \right) e^{-ity} dt = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left( n \sum_{l=1}^{\infty} \frac{k_l^{[Y]} - k_l^{[Z]}}{n^{\frac{l}{2}} l!} (it)^l \right) \exp \left( n \sum_{l=1}^{\infty} \frac{k_l^{[Z]}}{n^{\frac{l}{2}} l!} (it)^l \right) e^{-ity} dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left( n \sum_{l=1}^{\infty} \frac{k_l^{[Y]} - k_l^{[Z]}}{n^{\frac{l}{2}} l!} (it)^l \right) e^{-ity} E \left[ e^{it \frac{Z}{\sqrt{n}}} \right]^n dt, \end{aligned} \quad (31)$$

Empleamos ahora el operador derivada (18) como hemos hecho en repetidas ocasiones anteriormente,

$$f_{s_n}(y) = \exp \left( n \sum_{l=1}^{\infty} \frac{k_l^{[Y]} - k_l^{[Z]}}{n^{\frac{l}{2}} l!} (-1)^l D_y^l \right) \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} E \left[ e^{it \frac{Z}{\sqrt{n}}} \right]^n dt, \quad (32)$$

Llegados a este punto definimos una nueva variable  $W_n$  que tendrá como pdf a  $g_{W_n}(w)$ ,

$$W_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n Z_j$$

donde  $Z_1 = Z$  y las variables aleatorias  $\{Z_j\}_1^n$  son i.i.d.

Recordemos que al final de este desarrollo tomaremos  $n = 1$  para así obtener la serie de Edgeworth generalizada correspondiente a la pdf de  $Y$ . Notemos por tanto, que cuando  $n$  sea igual a 1 entonces  $W_1 = Z$  y  $g_{W_1}(w) = g_Z(w)$ .

Continuamos en (32) introduciendo esta nueva variable:

$$\begin{aligned} f_{S_n}(y) &= \exp \left( n \sum_{l=1}^{\infty} \frac{k_l^{[Y]} - k_l^{[Z]}}{n^{\frac{l}{2}} l!} (-1)^l D_y^l \right) \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} E(e^{itS_n}) dt \\ &= \exp \left( n \sum_{l=1}^{\infty} \frac{k_l^{[Y]} - k_l^{[Z]}}{n^{\frac{l}{2}} l!} (-1)^l D_y^l \right) g_{W_n}(y). \end{aligned} \quad (33)$$

Para terminar faltaría tomar el desarrollo de Taylor de la exponencial, ordenar la serie en función de las potencias de  $n$  y finalmente tomar  $n = 1$ . Este proceso se ve simplificado si la media y la varianza de  $Z$  coinciden con las de  $Y$  ya que entonces  $k_1^{[Y]} - k_1^{[Z]} = 0$  y  $k_2^{[Y]} - k_2^{[Z]} = 0$ . Dado que queremos aproximar la pdf de  $Y$  mediante la de  $Z$  es habitual ver este tipo de aproximaciones, por lo tanto continuaremos el proceso suponiendo que ambas variables coinciden en media y varianza.

Podemos eliminar de la sumatoria de (33) los dos primeros coeficientes, comenzando con  $l = 3$ ,

$$f_{S_n}(y) = \exp \left( n \sum_{l=3}^{\infty} \frac{k_l^{[Y]} - k_l^{[Z]}}{n^{\frac{l}{2}} l!} (-1)^l D_y^l \right) g_{W_n}(y). \quad (34)$$



Desarrollamos la exponencial y reordenamos según las potencias de  $n$ :

$$\begin{aligned}
f_{s_n}(y) &= g_{w_n}(y) + \\
&+ \frac{1}{\sqrt{n}} \left( \frac{k_3^{[Y]} - k_3^{[Z]}}{3!} (-1)^3 D_y^3 g_{w_n}(y) \right) + \\
&+ \frac{1}{\sqrt{n}^2} \left( \frac{k_4^{[Y]} - k_4^{[Z]}}{4!} (-1)^4 D_y^4 g_{w_n}(y) + \frac{(k_3^{[Y]} - k_3^{[Z]})^2}{2!3!3!} (-1)^6 D_y^6 g_{w_n}(y) \right) + \\
&+ \dots,
\end{aligned}$$

finalmente, tomando  $n = 1$ , la serie con los primeros términos es:

$$\begin{aligned}
f_Y(y) &= g_Z(y) + \\
&+ \left( -\frac{k_3^{[Y]} - k_3^{[Z]}}{3!} D_y^3 g_Z(y) \right) + \\
&+ \left( \frac{k_4^{[Y]} - k_4^{[Z]}}{4!} D_y^4 g_Z(y) + 10 \frac{(k_3^{[Y]} - k_3^{[Z]})^2}{6!} D_y^6 g_Z(y) \right) + \\
&+ \dots
\end{aligned} \tag{35}$$

### 3.4. Modificación de las series de Gram-Charlier y Edgeworth

A modo de profundización sobre las series estudiadas, vamos a realizar un cambio en su expresión. Se podrá comprobar en los experimentos realizados que esta aportación produce una mejora en los resultados.

Para la modificación en ambas series consideraremos  $Y$  como la variable a aproximar con  $f_Y$  como pdf. Necesitaremos también una variable que denotaremos como  $Z$  con pdf  $g_Z$  que tenga fórmula explícita conocida, y tal que los dos primeros momentos de  $Z$  coincidan con los de  $Y$ , o lo que es lo mismo,  $\mu^{[Y]} = \mu^{[Z]}$  y  $\sigma^{[Y]} = \sigma^{[Z]}$ .

#### 3.4.1. Modificación de la serie de Gram-Charlier

Empleando la serie de Gram-Charlier (20) tenemos que la pdf  $g_Z(z)$  se puede expresar como:

$$g_Z(z) = \phi(z, \mu^{[Z]}, \sigma^{[Z]}) \left[ 1 + \frac{k_3^{[Z]}}{3!\sigma^{[Z]3}} H_3(\hat{z}) + \frac{k_4^{[Z]}}{4!\sigma^{[Z]4}} H_4(\hat{z}) + \dots \right], \quad (36)$$

donde  $\hat{z} = \left( \frac{z - \mu^{[Z]}}{\sigma^{[Z]}} \right)$ .

Despejando  $\phi(z, \mu^{[Z]}, \sigma^{[Z]})$  de esta ecuación (36) se obtiene la pdf de una normal expresada en términos de la pdf de  $Z$ :

$$\phi(z, \mu^{[Z]}, \sigma^{[Z]}) = g_Z(z) \left[ 1 + \frac{k_3^{[Z]}}{3!\sigma^{[Z]3}} H_3(\hat{z}) + \frac{k_4^{[Z]}}{4!\sigma^{[Z]4}} H_4(\hat{z}) + \dots \right]^{-1}. \quad (37)$$

De igual manera que hicimos para  $g_Z$ , podemos obtener la pdf de  $Y$  como una serie:

$$f_Y(y) = \phi(y, \mu^{[Y]}, \sigma^{[Y]}) \left[ 1 + \frac{k_3^{[Y]}}{3!\sigma^{[Y]3}} H_3(\hat{y}) + \frac{k_4^{[Y]}}{4!\sigma^{[Y]4}} H_4(\hat{y}) + \dots \right], \quad (38)$$

con  $\hat{y} = \left( \frac{y - \mu^{[Y]}}{\sigma^{[Y]}} \right)$ .

Debido a la variable  $Z$  que hemos tomado, tenemos  $\mu^{[Z]} = \mu^{[Y]}$  y  $\sigma^{[Z]} = \sigma^{[Y]}$ , por lo tanto, si en la ecuación (37) evaluamos en  $y$  podremos sustituir en (38) la función  $\phi(y, \mu^{[Y]}, \sigma^{[Y]})$ , y obtendremos así otra forma de expresar la pdf de  $Y$ .

El resultado es:

$$f_Y(y) = g_Z(y) \frac{\left( 1 + \frac{k_3^{[Y]}}{3!\sigma^{[Y]3}} H_3(\hat{y}) + \frac{k_4^{[Y]}}{4!\sigma^{[Y]4}} H_4(\hat{y}) + \dots \right)}{\left( 1 + \frac{k_3^{[Z]}}{3!\sigma^{[Y]3}} H_3(\hat{y}) + \frac{k_4^{[Z]}}{4!\sigma^{[Y]4}} H_4(\hat{y}) + \dots \right)}, \quad (39)$$

donde  $\hat{y} = \left( \frac{y - \mu^{[Y]}}{\sigma^{[Y]}} \right)$  y  $g_z(y)$  toma su fórmula conocida.

### 3.4.2. Modificación de la serie de Edgeworth

El procedimiento a seguir para esta serie es igual al anterior, cambiando eso sí, las expresiones de la serie de Gram-Charlier por las de Edgeworth. De esta forma empleando (??) tendremos

$$g_z(z) = \phi(z, \mu^{[Z]}, \sigma^{[Z]}) \left[ 1 + \left( \frac{k_3^{[Z]}}{3!\sigma^{[Z]3}} H_3(\hat{z}) \right) + \right. \\ \left. + \left( \frac{k_4^{[Z]}}{4!\sigma^{[Z]4}} H_4(\hat{z}) + \frac{10k_3^{[Z]2}}{6!\sigma^{[Z]6}} H_6(\hat{z}) \right) + \dots \right], \quad (40)$$

donde  $\hat{z} = \left( \frac{z - \mu^{[Z]}}{\sigma^{[Z]}} \right)$ .

Despejamos  $\phi(z, \mu^{[Z]}, \sigma^{[Z]})$ ,

$$\phi(z, \mu^{[Z]}, \sigma^{[Z]}) = g_z(z) \left[ 1 + \left( \frac{k_3^{[Z]}}{3!\sigma^{[Z]3}} H_3(\hat{z}) \right) + \right. \\ \left. + \left( \frac{k_4^{[Z]}}{4!\sigma^{[Z]4}} H_4(\hat{z}) + \frac{10k_3^{[Z]2}}{6!\sigma^{[Z]6}} H_6(\hat{z}) \right) + \dots \right]^{-1}. \quad (41)$$

Por otro lado tenemos  $f_Y$  como una serie de Edgeworth:

$$f_Y(y) = \phi(y, \mu^{[Y]}, \sigma^{[Y]}) \left[ 1 + \left( \frac{k_3^{[Y]}}{3!\sigma^{[Y]3}} H_3(\hat{y}) \right) + \right. \\ \left. + \left( \frac{k_4^{[Y]}}{4!\sigma^{[Y]4}} H_4(\hat{y}) + \frac{10k_3^{[Y]2}}{6!\sigma^{[Y]6}} H_6(\hat{y}) \right) + \dots \right], \quad (42)$$

donde  $\hat{y} = \left( \frac{y - \mu^{[Y]}}{\sigma^{[Y]}} \right)$ .

Recordando que los dos primeros momentos de  $Y$  y  $Z$  coinciden, evaluando la función (41) en  $y$  y sustituyendo en la ecuación anterior, tenemos la modificación deseada:

$$f_Y(y) = g_Z(y) \frac{\left(1 + \left(\frac{k_3^{[Y]}}{3!\sigma^{[Y]3}} H_3(\hat{y})\right) + \left(\frac{k_4^{[Y]}}{4!\sigma^{[Y]4}} H_4(\hat{y}) + 10 \frac{k_3^{[Y]2}}{6!\sigma^{[Y]6}} H_6(\hat{y})\right) + \dots\right)}{\left(1 + \left(\frac{k_3^{[Z]}}{3!\sigma^{[Y]3}} H_3(\hat{y})\right) + \left(\frac{k_4^{[Z]}}{4!\sigma^{[Y]4}} H_4(\hat{y}) + 10 \frac{k_3^{[Z]2}}{6!\sigma^{[Y]6}} H_6(\hat{y})\right) + \dots\right)}, \quad (43)$$

donde  $\hat{y} = \left(\frac{y - \mu^{[Y]}}{\sigma^{[Y]}}\right)$  y  $g_Z(y)$  toma su definición explícita conocida.

### 3.5. Transformada de Fourier para pdf de sumas de variables aleatorias

En esta sección se muestra como expresar una pdf mediante transformadas de Fourier directa e inversa, las cuales están definidas en el anexo C.

Sea  $\{X_j\}_1^N$  un conjunto de v.a.i.i.d, con pdf  $f_X$ , definimos

$$Y = \sum_{j=1}^N X_j. \quad (44)$$

Expresamos la pdf de  $Y$  en función de  $f_X$  y la Delta de Dirac:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(x_1) \dots f_X(x_N) \delta\left(y - \sum_{j=1}^N X_j\right) dx_1 \dots dx_N \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \delta\left(y - \sum_{j=1}^N X_j\right) \left(\prod_{j=1}^N f_X(x_j)\right) dx_1 \dots dx_N, \end{aligned} \quad (45)$$

sustituimos ahora el símbolo de la Delta de Dirac por su definición (82),

$$\begin{aligned} f_Y(y) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{iky} \exp\left(-ik \sum_{j=1}^N x_j\right) \left(\prod_{j=1}^N f_X(x_j)\right) dk \, dx_1 \dots dx_N \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{iky} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left(\prod_{j=1}^N e^{(-ikx_j)} f_X(x_j)\right) dx_1 \dots dx_N \, dk \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{iky} \left(\prod_{j=1}^N \int_{-\infty}^{\infty} e^{(-ikx_j)} f_X(x_j) dx_j\right) dk \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{iky} \left(\int_{-\infty}^{\infty} e^{(-ikx_j)} f_X(x_j) dx_j\right)^N dk, \end{aligned} \quad (46)$$

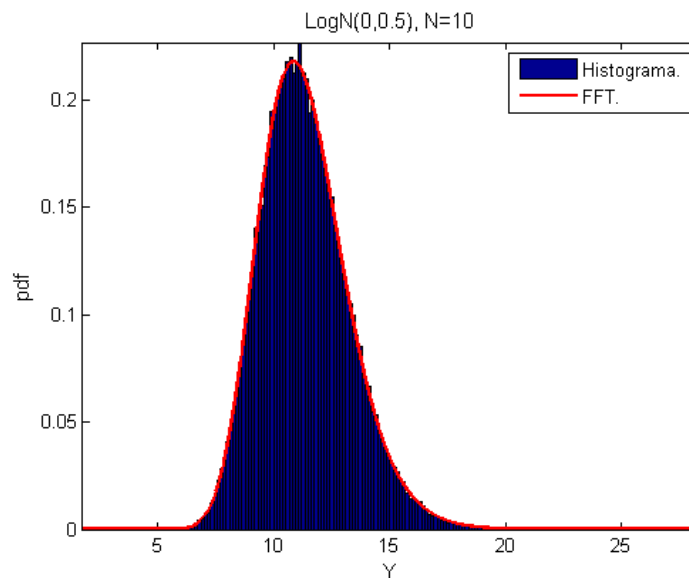
observemos que podemos sustituir por la transformada de Fourier directa (79),

$$f_Y(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{iky} \left(\hat{f}_X(k)\right)^N dk. \quad (47)$$

Y por (79) vemos que  $f_Y(y)$  es la transformada de Fourier inversa de  $(\hat{f}_X(k))^N$ .

El inconveniente de esta solución se encuentra cuando se lleva a la práctica. A pesar de que algunas herramientas de software matemático como MATLAB disponen de transformadas de Fourier rápidas, que calculan la transformada a partir de un conjunto limitado de puntos, su coste computacional es elevado.

En el caso particular de la resolución de (47), empleando las funciones de transformadas rápidas de Fourier de MATLAB y ordenador Acer Aspire 5730Z con memoria RAM 4096 MB, se presenta el problema de que al tomar  $N$  superior a 70 aproximadamente no se obtiene solución. Para valores más pequeños de  $N$  se necesita que el conjunto de puntos tome valores tanto de la variable individual  $X$  como de la suma de variables, ya que en primer lugar se calcula la transformada directa de  $f_X(x)$  y posteriormente la transformada inversa sobre  $(\hat{f}_X(k))^N$ .



**Figura 2:** Ejemplo de histograma y pdf calculada con transformadas de Fourier rápidas, para una suma de 10 variables lognormales con  $\mu = 0$  y  $\sigma = 0.5$ .

El número de puntos del soporte debe ser además denso para que la solución sea la más ajustada posible, es esto lo que hace que tenga un coste computacional muy alto y por ello se exploren otras opciones de aproximación, como las series estudiadas anteriormente.

## 4. Librería desarrollada

Las distintas series presentadas en la sección anterior han sido implementadas con el fin de crear una librería completa que permita obtener resultados numéricos.

Para la creación de dicha librería y los programas utilizados en la fase de pruebas se ha empleado la herramienta de software MATLAB, ya que se adecua a las necesidades del entorno matemático en el que se enmarca este trabajo.

A continuación se presenta la documentación de la librería. Donde se incluyen las funciones correspondientes al cálculo de las distintas modalidades de las series de Gram-Charlier y Edgeworth, así como las funciones para el cálculo de la pdf de una suma de lognormales mediante el método de Monte Carlo y con la transformada de Fourier.

### 4.1. Documentación de las funciones de la librería

#### Serie de Gram-Charlier

Cabecera: `Gram_Charlier(f, h, k, sg)`

Descripción: calcula la serie de Gram-Charlier hasta orden 5 para una variable aleatoria.

Entradas:

- f: pdf gaussiana evaluada en distintos valores que toma la variable aleatoria.
- h: 5 primeros polinomios de hermite evaluados en la variable aleatoria estandarizada.
- k: 5 primeros cumulantes de la variable aleatoria.
- sg: desviación típica de la variable aleatoria.

Salida:

- Matriz donde la primera, segunda, tercera y cuarta fila corresponden a orden 0, 3, 4 y 5 respectivamente.

#### Serie de Edgeworth

Cabecera: `Edgeworth(f, h, k, sg)`

Descripción: calcula la serie de Edgeworth hasta orden 5 para una variable aleatoria.

Entradas:

- f: pdf gaussiana evaluada en distintos valores que toma la variable aleatoria.
- h: 9 primeros polinomios de hermite evaluados en la variable aleatoria estandarizada.
- k: 5 primeros cumulantes de la variable aleatoria.
- sg: desviación típica de la variable aleatoria.

Salida:

- Matriz donde la primera, segunda, tercera y cuarta fila corresponden a orden 0, 3, 4 y 5 respectivamente.

**Serie de Gram-Charlier generalizada**

Cabecera: `general_GC(f, deriv_f, k_Y, k_Z)`

Descripción: calcula la serie de Gram-Charlier generalizada hasta orden 5 para una variable aleatoria  $Y$ , utilizando como variable auxiliar  $Z$ .  $Y$  y  $Z$  coinciden en media y varianza.

Entradas:

- `f`: pdf de  $Z$ .
- `deriv_f`: 5 primeras derivadas de  $f$ .
- `k_Y`: 5 primeros cumulantes de  $Y$ .
- `k_Z`: 5 primeros cumulantes de  $Z$ .

Salida:

- Matriz donde la primera, segunda, tercera y cuarta fila corresponden a orden 0, 3, 4 y 5 respectivamente.

**Serie de Edgeworth generalizada**

Cabecera: `general_Edg(f, deriv_f, k_Y, k_Z)`

Descripción: calcula la serie de Edgeworth generalizada hasta orden 5 para una variable aleatoria  $Y$ , utilizando como variable auxiliar  $Z$ .  $Y$  y  $Z$  coinciden en media y varianza.

Entradas:

- `f`: pdf de  $Z$ .
- `deriv_f`: 9 primeras derivadas de  $f$ .
- `k_Y`: 5 primeros cumulantes de  $Y$ .
- `k_Z`: 5 primeros cumulantes de  $Z$ .

Salida:

- Matriz donde la primera, segunda, tercera y cuarta fila corresponden a orden 0, 3, 4 y 5 respectivamente.

**Serie de Gram-Charlier modificada**

Cabecera: `modified_GC(f, h, k_Y, k_Z, sg)`

Descripción: calcula la serie de Gram-Charlier modificada hasta orden 5 para una variable aleatoria  $Y$ , utilizando como variable auxiliar  $Z$ .  $Y$  y  $Z$  coinciden en media y varianza.

Entradas:

- `f`: pdf de  $Z$ .
- `h`: 5 primeros polinomios de hermite evaluados en la variable  $Y$  estandarizada.
- `k_Y`: 5 primeros cumulantes de  $Y$ .
- `k_Z`: 5 primeros cumulantes de  $Z$ .
- `sg`: desviación típica de  $Y$ .

Salida:

- Matriz donde la primera, segunda, tercera y cuarta fila corresponden a orden 0, 3, 4 y 5 respectivamente.



### Serie de Edgeworth modificada

Cabecera: `modified_Edg(f, h, k_Y, k_Z, sg)`

Descripción: calcula la serie de Edgeworth modificada hasta orden 5 para una variable aleatoria  $Y$ , utilizando como variable auxiliar  $Z$ .  $Y$  y  $Z$  coinciden en media y varianza.

Entradas:

- $f$ : pdf de  $Z$ .
- $h$ : 9 primeros polinomios de hermite evaluados en la variable  $Y$  estandarizada.
- $k_Y$ : 5 primeros cumulantes de  $Y$ .
- $k_Z$ : 5 primeros cumulantes de  $Z$ .
- $sg$ : desviación típica de  $Y$ .

Salida:

- Matriz donde la primera, segunda, tercera y cuarta fila corresponden a orden 0, 3, 4 y 5 respectivamente.

### Método de Monte Carlo

Cabecera: `N_logn_Monte_Carlo(N, mu, sg)`

Descripción: calcula muestras de una suma de  $N$  lognormales,  $X_1 + \dots + X_N$ , donde  $X_1, \dots, X_N$  son v.a.i.i.d.

Entradas:

- $N$ : número de variables en la suma.
- $mu$ : media de  $\log(X_1)$ .
- $sg$ : desviación típica de  $\log(X_1)$ .

Salida:

- Array de valores tomados por la suma de variables aleatorias.

### Transformada de Fourier para pdf

Cabecera: `N_lognpdf_fft(N, x, mu, sg)`

Descripción: empleando la transformada de Fourier rápida calcula la pdf de una suma de  $N$  lognormales,  $X_1 + \dots + X_N$ , donde  $X_1, \dots, X_N$  son v.a.i.i.d.

Entradas:

- $N$ : número de variables en la suma.
- $x$ : array puntos sobre los que calcular la pdf, su tamaño debe ser potencia de 2 y debe dar soporte tanto a la pdf de los valores individuales de la suma como a la pdf de los valores agregados.
- $mu$ : media de  $\log(X_1)$ .
- $sg$ : desviación típica de  $\log(X_1)$ .

Salida:

- pdf evaluada en el conjunto de puntos  $x$ .



## 5. Experimentos numéricos

Todas las pruebas han sido realizadas sobre la distribución de una suma de lognormales, ya que éstas tienen todos los momentos definidos y en consecuencia los cumulantes. Otro factor decisivo a la hora de realizar los experimentos sobre lognormales es el importante uso que tienen en aplicaciones de interés práctico, como las telecomunicaciones y las finanzas.

A continuación se establecen las bases comunes a todas las pruebas realizadas.

Utilizaremos la nomenclatura  $N(\mu, \sigma)$  para referirnos a una variable con distribución normal de media  $\mu$  y varianza  $\sigma^2$ . La función de densidad de probabilidad de una normal es:

$$\phi(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (48)$$

y su función generadora de momentos es:

$$M(t) = e^{t\mu + t^2\sigma^2/2}. \quad (49)$$

Una variable  $X$  tendrá una distribución lognormal con parámetros  $\mu$  y  $\sigma$  cuando

$$\log(X) \sim N(\mu, \sigma),$$

diremos entonces que

$$X \sim \text{Log}N(\mu, \sigma)$$

donde  $\mu$  será la media de  $\log(X)$  y  $\sigma$  su desviación típica.

La función de densidad de una lognormal es:

$$pdf(x, \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}. \quad (50)$$

y dado que  $\log(X)$  es una normal, podemos encontrar fácilmente los momentos no centrales utilizando (49),

$$E(X^t) = E(e^{\log(X)t}) = e^{t\mu + t^2\sigma^2/2}. \quad (51)$$

Una vez vistos estos conceptos podemos comenzar a definir las variables utilizadas para la realización de las distintas pruebas.

La variable de la que aproximar su función de densidad de probabilidad es  $Y$ , definida como una suma de  $N$  variables,

$$Y = \sum_{j=1}^N X_j, \quad (52)$$

donde  $\{X_j\}_1^N$  son v.a.i.i.d. con distribución lognormal, de parámetros media  $\mu$  y varianza  $\sigma^2$ ,

$$X_j \sim \text{LogN}(\mu, \sigma). \quad (53)$$

La media y desviación típica de  $Y$  se denotarán como  $\mu^{[Y]}$  y  $\sigma^{[Y]}$  respectivamente,

$$\begin{aligned} \mu^{[Y]} &= E(Y) = E\left(\sum_{j=1}^N X_j\right) = NE(X_1) \\ \sigma^{[Y]} &= V(Y) = V\left(\sum_{j=1}^N X_j\right) = NV(X_1) \end{aligned}$$

En las series estudiadas en las anteriores secciones siempre era necesario el uso de los cumulantes de la variable a aproximar, en este caso los de  $Y$ . Se puede utilizar la propiedad aditiva de los cumulantes para expresar el cumulante de  $Y$  en función del cumulante de  $X_1$  (70). De esta forma únicamente habrá que calcular los cumulantes de  $X_1$ ,

$$k_n^{[Y]} = Nk_n^{[X_1]}. \quad (54)$$

Los cumulantes se pueden calcular mediante los momentos y cumulantes de orden inferior (71), por lo que calculando de forma recursiva basta con conocer los momentos no centrales de  $X_1$ , que pueden ser obtenidos utilizando (51).

Como pdf exacta de  $Y$  con la que comparar las aproximaciones se tomará la solución obtenida mediante las transformadas de Fourier (47).

A continuación veremos tres versiones diferentes para aproximar una suma de variables aleatorias empleando las distintas series desarrolladas anteriormente. Para cada una de las versiones se realizarán una serie de pruebas, en las que se fija el parámetro  $\mu = 0$  y se varía el valor del parámetro  $\sigma$ . Ya que  $\sigma$  determina cómo de pesada es la cola de la distribución, a mayores valores de  $\sigma$  más pesada será la cola.

## 5.1. Versión 1

En esta primera versión consideraremos las series tanto de Gram-Charlier (20) como de Edgeworth (28) básicas, es decir, empleando como función de referencia la pdf gaussiana.

De esta manera tenemos las siguientes series:

Versión 1 Gram-Charlier:

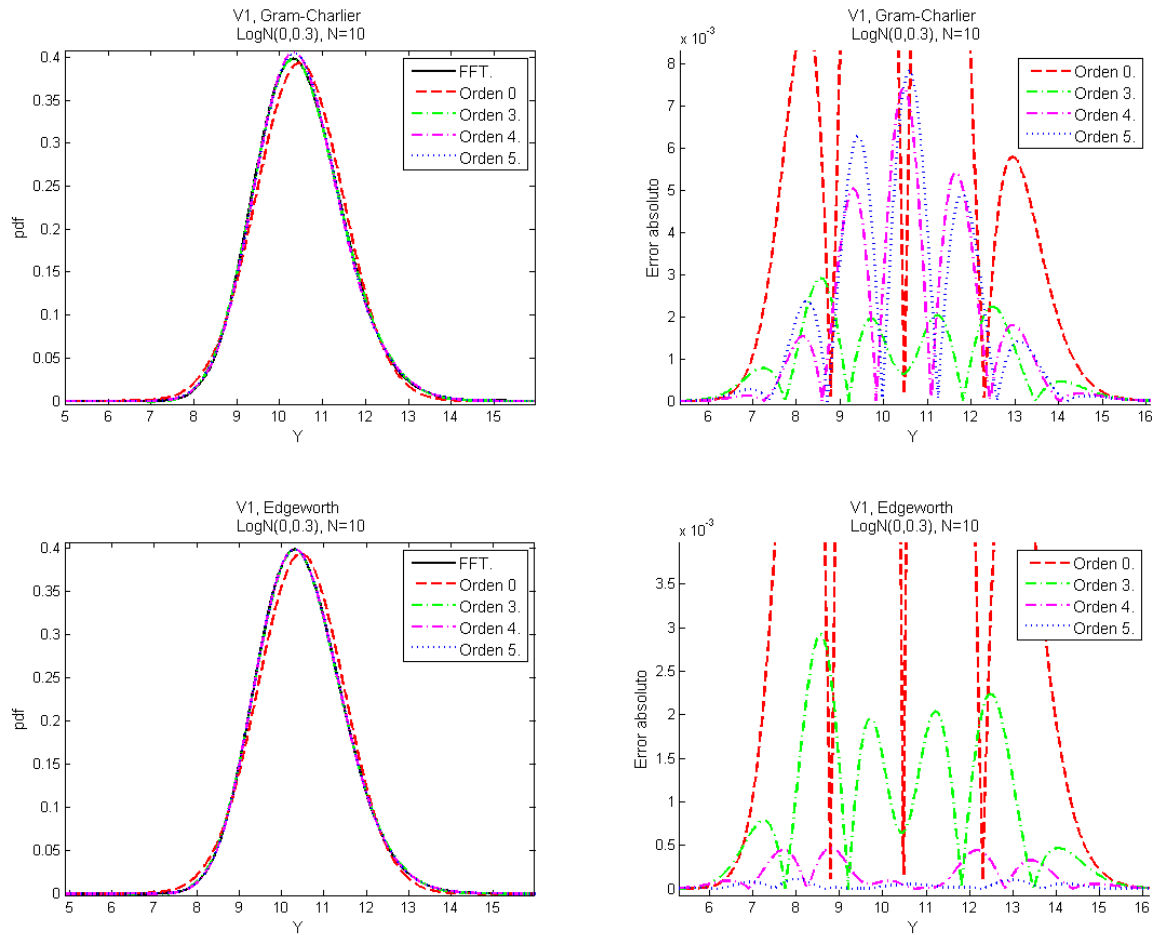
$$\begin{aligned}
 f_Y(y) = \phi(y, \mu^{[Y]}, \sigma^{[Y]}) & \left[ 1 + \right. \\
 & + \frac{k_3}{3! \sigma^{[Y]3}} H_3\left(\frac{y - \mu^{[Y]}}{\sigma^{[Y]}}\right) + \\
 & + \frac{k_4}{4! \sigma^{[Y]4}} H_4\left(\frac{y - \mu^{[Y]}}{\sigma^{[Y]}}\right) + \\
 & \left. + \frac{k_5}{5! \sigma^{[Y]5}} H_5\left(\frac{y - \mu^{[Y]}}{\sigma^{[Y]}}\right) + \dots \right],
 \end{aligned} \tag{55}$$

Versión 1 Edgeworth:

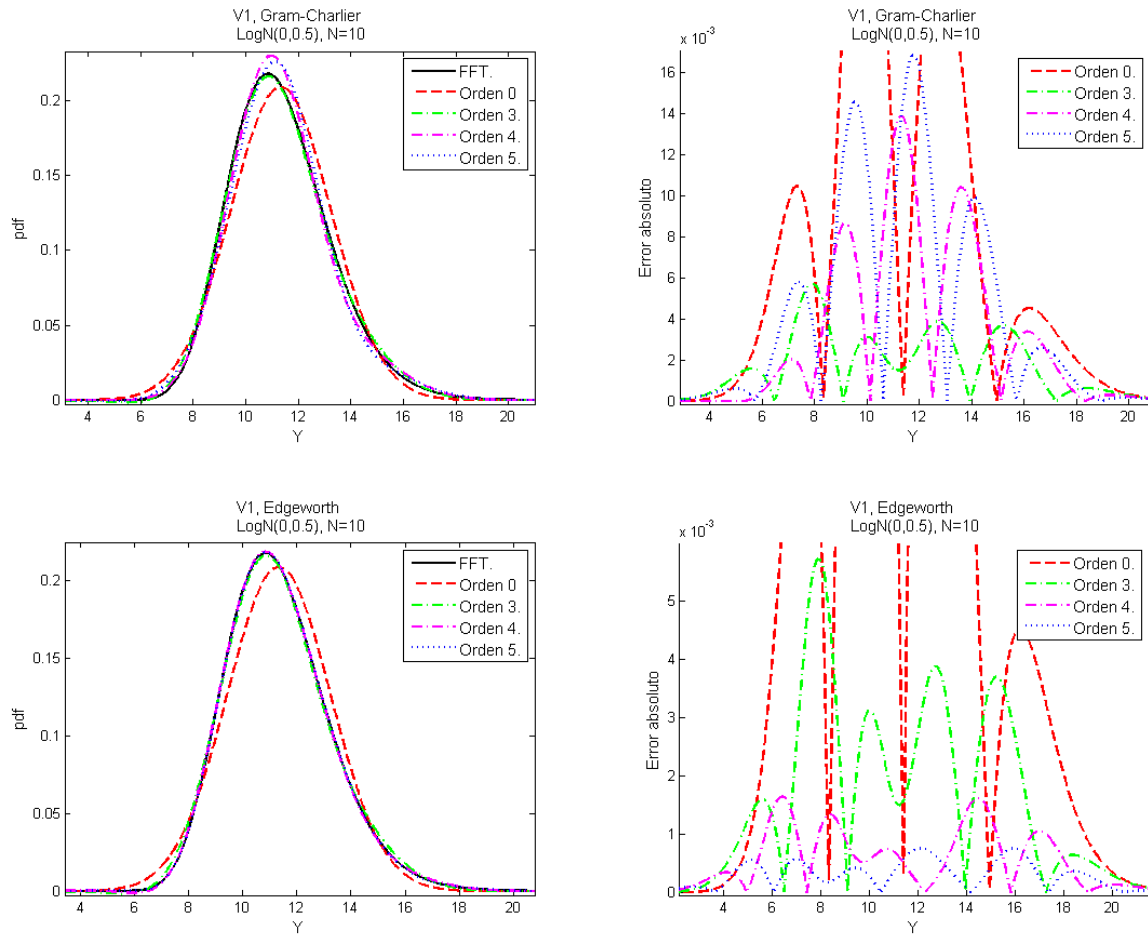
$$\begin{aligned}
 f_Y(y) = \phi(y, \mu^{[Y]}, \sigma^{[Y]}) & \left[ 1 + \right. \\
 & + \left( \frac{k_3}{3! \sigma^{[Y]3}} H_3\left(\frac{y - \mu^{[Y]}}{\sigma^{[Y]}}\right) \right) + \\
 & + \left( \frac{k_4}{4! \sigma^{[Y]4}} H_4\left(\frac{y - \mu^{[Y]}}{\sigma^{[Y]}}\right) + 10 \frac{k_3^2}{6!} H_6\left(\frac{y - \mu^{[Y]}}{\sigma^{[Y]}}\right) \right) + \\
 & + \left( \frac{k_5}{5! \sigma^{[Y]5}} H_5\left(\frac{y - \mu^{[Y]}}{\sigma^{[Y]}}\right) + 35 \frac{k_3 k_4}{7! \sigma^{[Y]7}} H_7\left(\frac{y - \mu^{[Y]}}{\sigma^{[Y]}}\right) + 280 \frac{k_3^3}{9! \sigma^{[Y]9}} H_9\left(\frac{y - \mu^{[Y]}}{\sigma^{[Y]}}\right) \right) + \\
 & \left. + \dots \right].
 \end{aligned} \tag{56}$$

### 5.1.1. Suma de 10 variables

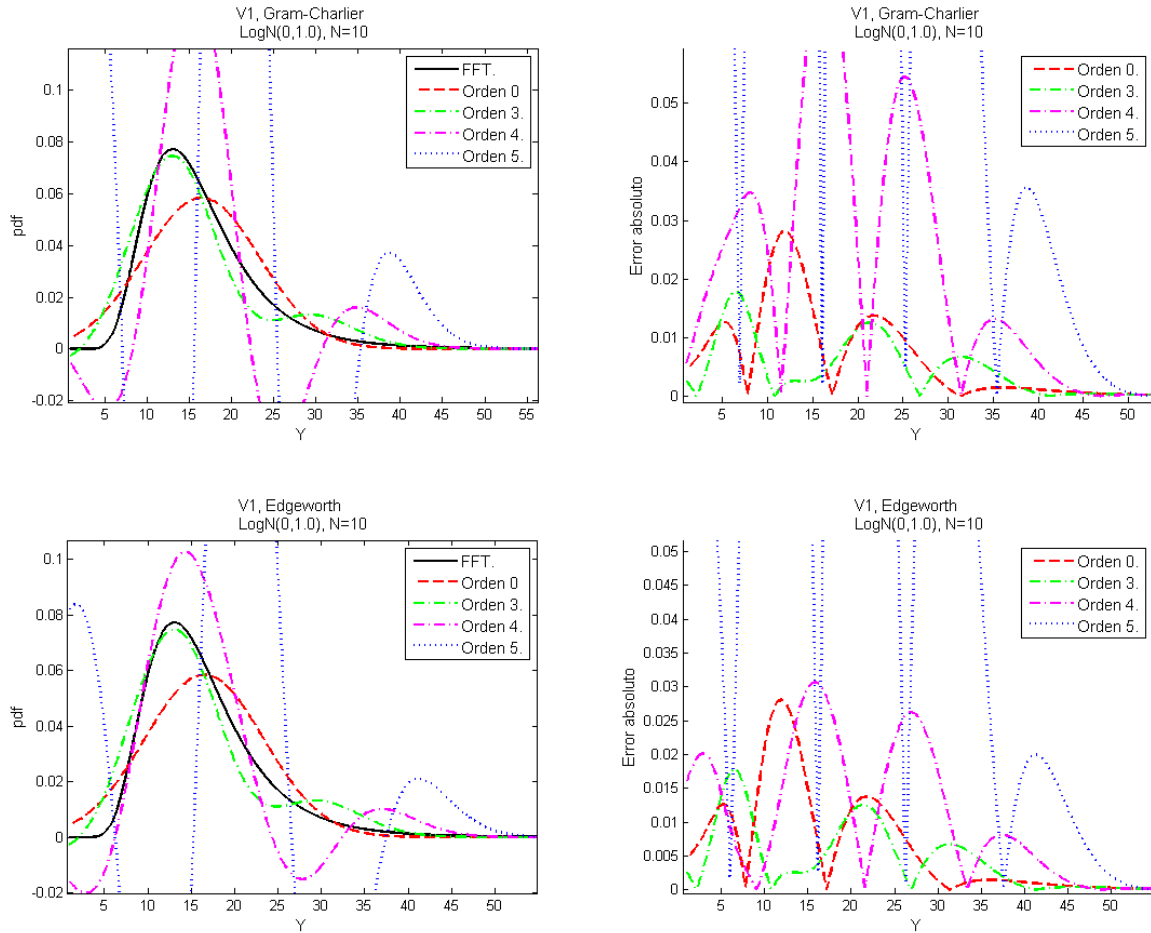
A continuación se presentan las gráficas correspondientes a la versión 1 cuando  $Y$  es una suma de diez variables,  $N = 10$ . Se fijará  $\mu = 0$  y se modificará el valor de  $\sigma$  para estudiar la evolución de las series respecto a este parámetro.



**Figura 3:** Aproximación a la pdf de una suma de 10 variables lognormales con  $\mu = 0$  y  $\sigma = 0.3$ . En la parte superior se emplea Gram-Charlier y en la inferior Edgeworth. De izquierda a derecha se encuentra la pdf y el error absoluto.



**Figura 4:** Aproximación a la pdf de una suma de 10 variables lognormales con  $\mu = 0$  y  $\sigma = 0.5$ . En la parte superior se emplea Gram-Charlier y en la inferior Edgeworth. De izquierda a derecha se encuentra la pdf y el error absoluto.



**Figura 5:** Aproximación a la pdf de una suma de 10 variables lognormales con  $\mu = 0$  y  $\sigma = 0.1$ . En la parte superior se emplea Gram-Charlier y en la inferior Edgeworth. De izquierda a derecha se encuentra la pdf y el error absoluto.

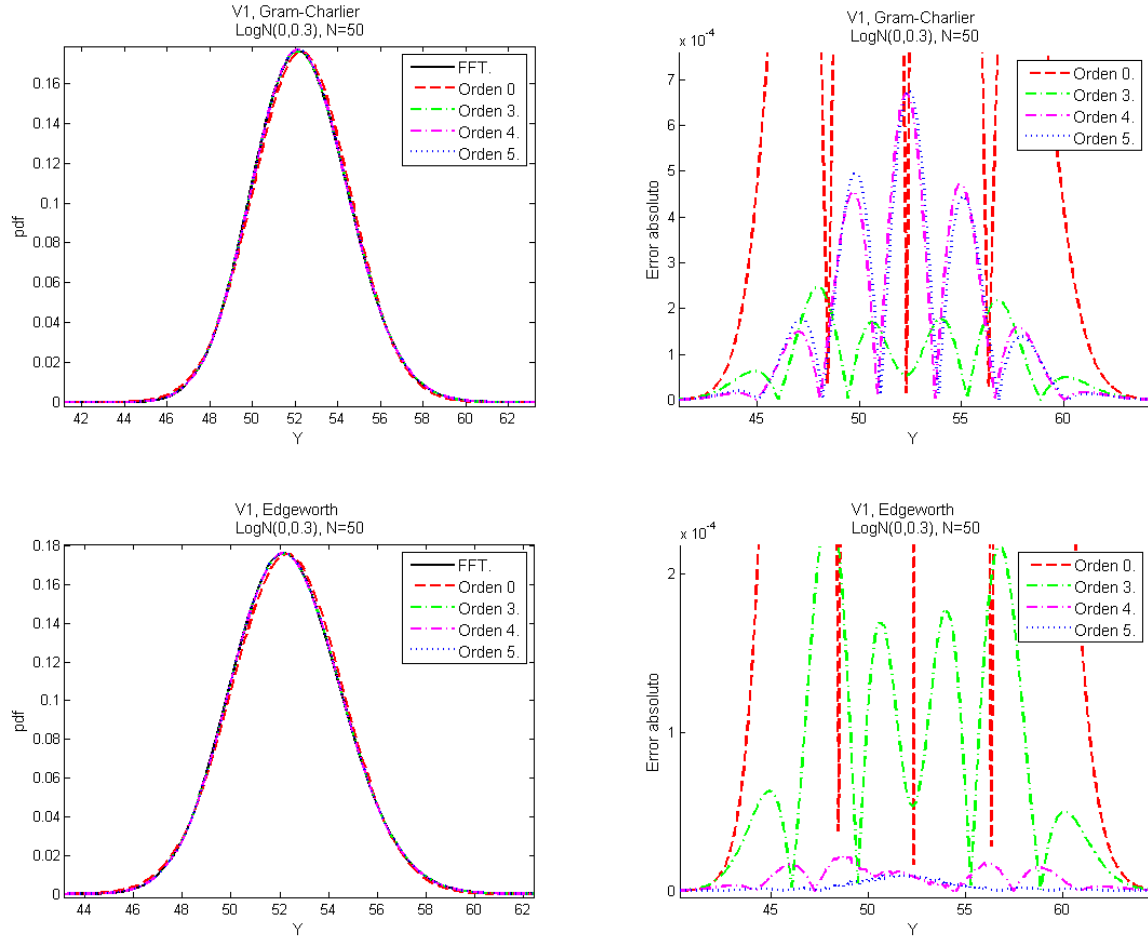
Observaciones sobre las figuras (3), (4) y (5) :

1. Cabe destacar que ninguna de las series truncadas obtiene una aproximación ajustada para valores de  $\sigma$  superiores a 0.7 aproximadamente.
2. Para valores de sigma pequeños ( $\sigma \leq 0.6$  aproximadamente) la serie de Edgeworth mejora las aproximaciones de orden 4 y 5 de Gram-Charlier.
3. En la serie de Gram-Charlier la mejor aproximación se obtiene con orden 3, mientras que en la serie de Edgeworth se obtiene con orden 5 para valores pequeños de  $\sigma$  y orden 3 para el resto.
4. Recordando las fórmulas empleadas (55) y (56) tenemos que orden 0 en ambos casos corresponde a la pdf gaussiana. En las gráficas correspondientes al error absoluto se observa como empeora al aumentar  $\sigma$ , esto se debe a que la pdf lognormal se aleja de su parecido inicial a la normal.

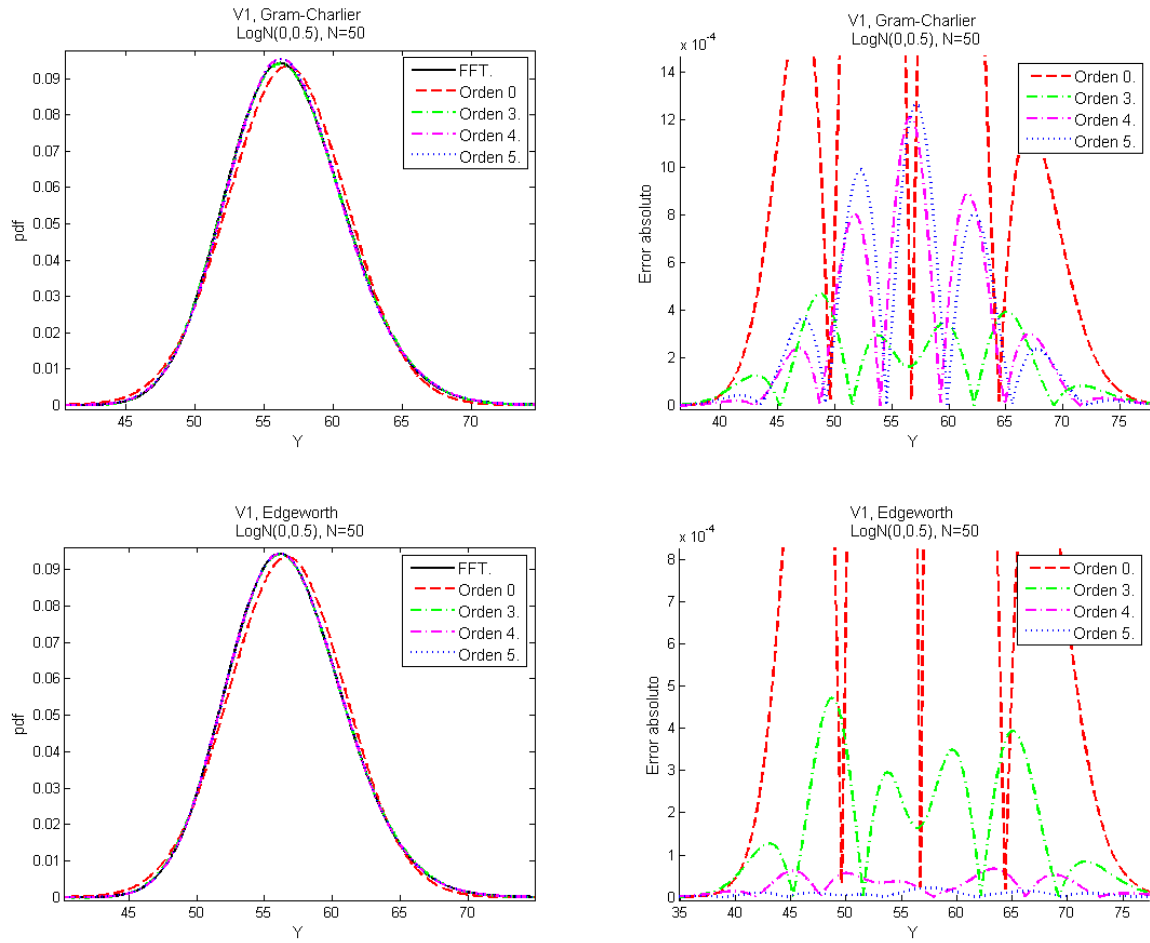


### 5.1.2. Suma de 50 variables

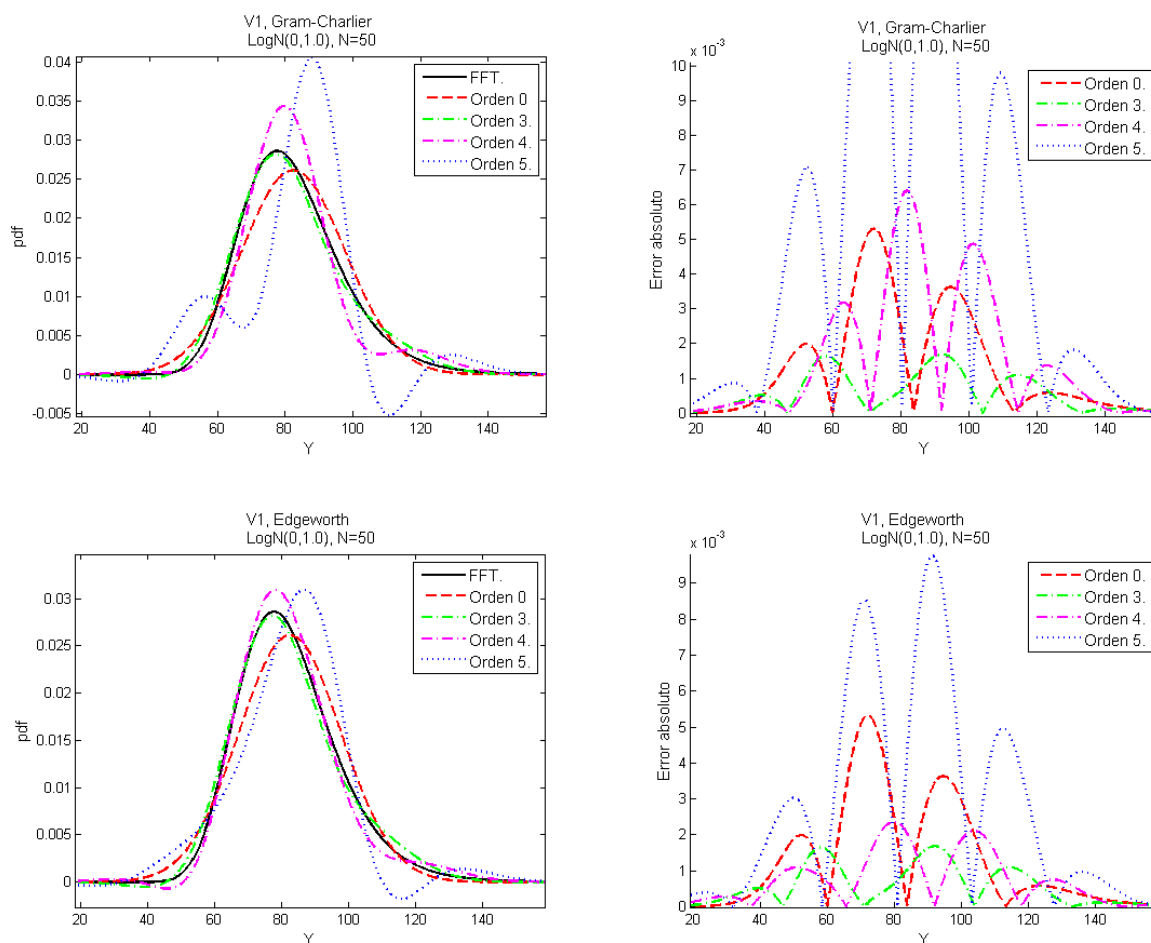
Resultados gráficos de la versión 1 con una suma de 50 variables para  $\mu = 0$  y  $\sigma$  tomando los valores 0.3, 0.5 y 1.



**Figura 6:** Aproximación a la pdf de una suma de 50 variables lognormales con  $\mu = 0$  y  $\sigma = 0.3$ . En la parte superior se emplea Gram-Charlier y en la inferior Edgeworth. De izquierda a derecha se encuentra la pdf y el error absoluto.



**Figura 7:** Aproximación a la pdf de una suma de 50 variables lognormales con  $\mu = 0$  y  $\sigma = 0.5$ . En la parte superior se emplea Gram-Charlier y en la inferior Edgeworth. De izquierda a derecha se encuentra la pdf y el error absoluto.



**Figura 8:** Aproximación a la pdf de una suma de 50 variables lognormales con  $\mu = 0$  y  $\sigma = 1$ . En la parte superior se emplea Gram-Charlier y en la inferior Edgeworth. De izquierda a derecha se encuentra la pdf y el error absoluto.

En estas simulaciones con  $N = 50$  se conserva la esencia de los resultados obtenidos para  $N = 10$ , es decir, Edgeworth sigue mejorando a Gram-Charlier en orden 4 y 5, y conforme aumenta el valor de  $\sigma$  las aproximaciones empeoran. También se mantiene que a partir de cierto  $\sigma$  se observa que en Edgeworth sus mejores truncaciones dejan de serlo.

Sin embargo, sí que encuentra un cambio significativo al aumentar el número de variables en la suma. El error se ve claramente reducido. Por ejemplo, tomando  $\sigma = 1$ , en la figura (5) la mejor aproximación corresponde a orden 3 teniendo un máximo de error mayor a 0.015 y comparando con la figura (8) el máximo error de orden 3 no supera a 0.002.

## 5.2. Versión 2

En esta versión utilizaremos las modificaciones realizadas sobre las series de Gram-Charlier y Edgeworth. En este caso, para poder emplearlas tomaremos la variable  $Z$  con distribución lognormal que aproxime a la suma de lognormales  $Y$ , haciendo coincidir los dos primeros momentos de ambas variables, es decir, utilizaremos el método de Fenton-Wilkinson para hallar los parámetros de la distribución lognormal de  $Z$ .

Tenemos pues,

$$Z \sim \text{LogN}(\mu', \sigma'),$$

donde, al ser  $Y$  suma de lognormales todas ellas con misma media ( $\mu$ ) y varianza ( $\sigma^2$ ),  $\mu'$  y  $\sigma'$  se calculan como:

$$\sigma'^2 = \log\left(\frac{e^\sigma - 1}{N} + 1\right)$$

$$\mu' = \log(Ne^\mu) + \frac{\sigma^2}{2} + \frac{\sigma'^2}{2}$$

Si  $g_z(y)$  es la pdf de  $Z$ , recordando (50):

$$g_z(z) = \frac{1}{z\sigma'\sqrt{2\pi}} e^{-\frac{(\ln(z)-\mu')^2}{2\sigma'^2}}. \quad (57)$$

Basta ahora con tomar la ecuación de la serie de Gram-Charlier modificada (39):

Versión 2 Gram-Charlier:

$$f_Y(y) = g_Z(y) \frac{\left(1 + \frac{k_3^{[Y]}}{3!\sigma^{[Y]3}} H_3(\hat{y}) + \frac{k_4^{[Y]}}{4!\sigma^{[Y]4}} H_4(\hat{y}) + \dots\right)}{\left(1 + \frac{k_3^{[Z]}}{3!\sigma^{[Y]3}} H_3(\hat{y}) + \frac{k_4^{[Z]}}{4!\sigma^{[Y]4}} H_4(\hat{y}) + \dots\right)}, \quad (58)$$

donde  $\hat{y} = \left(\frac{y-\mu^{[Y]}}{\sigma^{[Y]}}\right)$  y  $g_Z(y)$  toma su definición de lognormal (50).

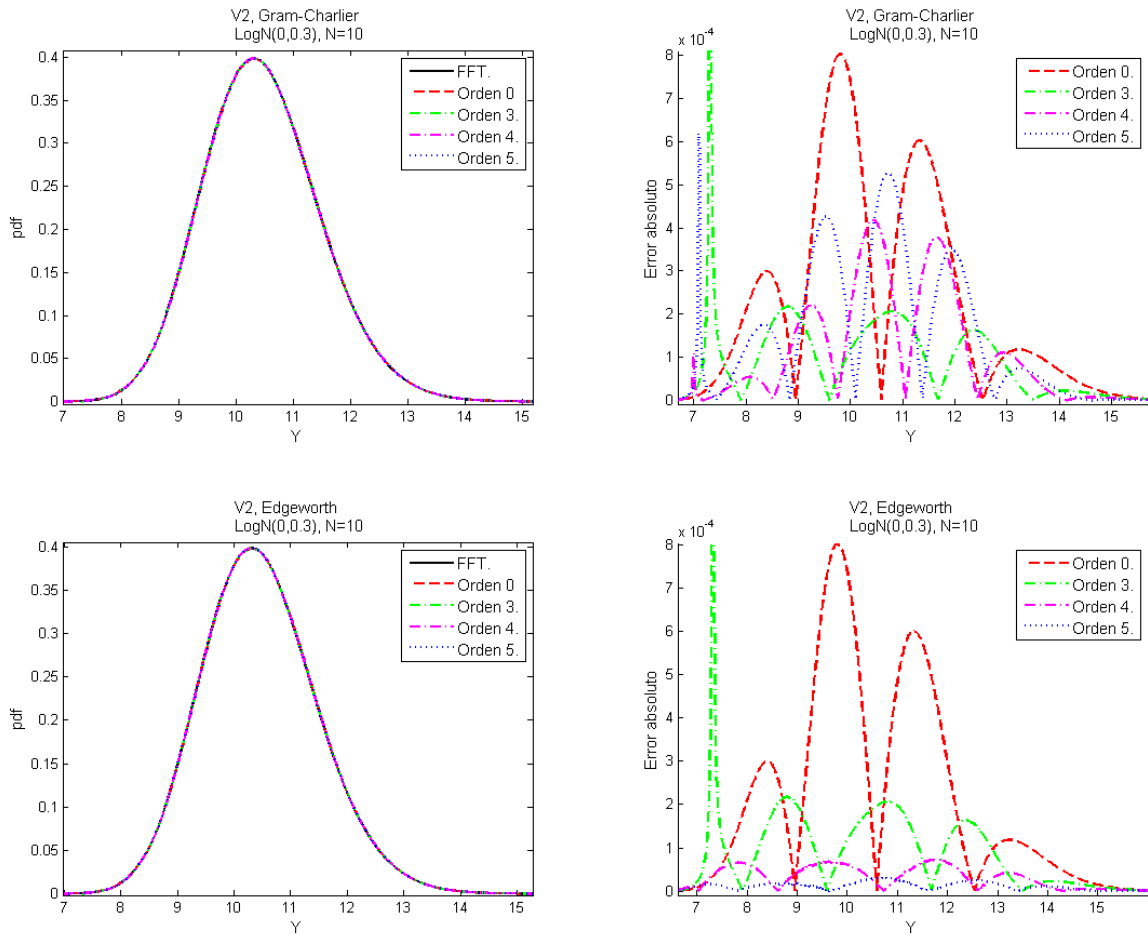
Utilizando la correspondiente fórmula de la serie de Edgeworth modificada obtendremos:

Versión 2 Edgeworth:

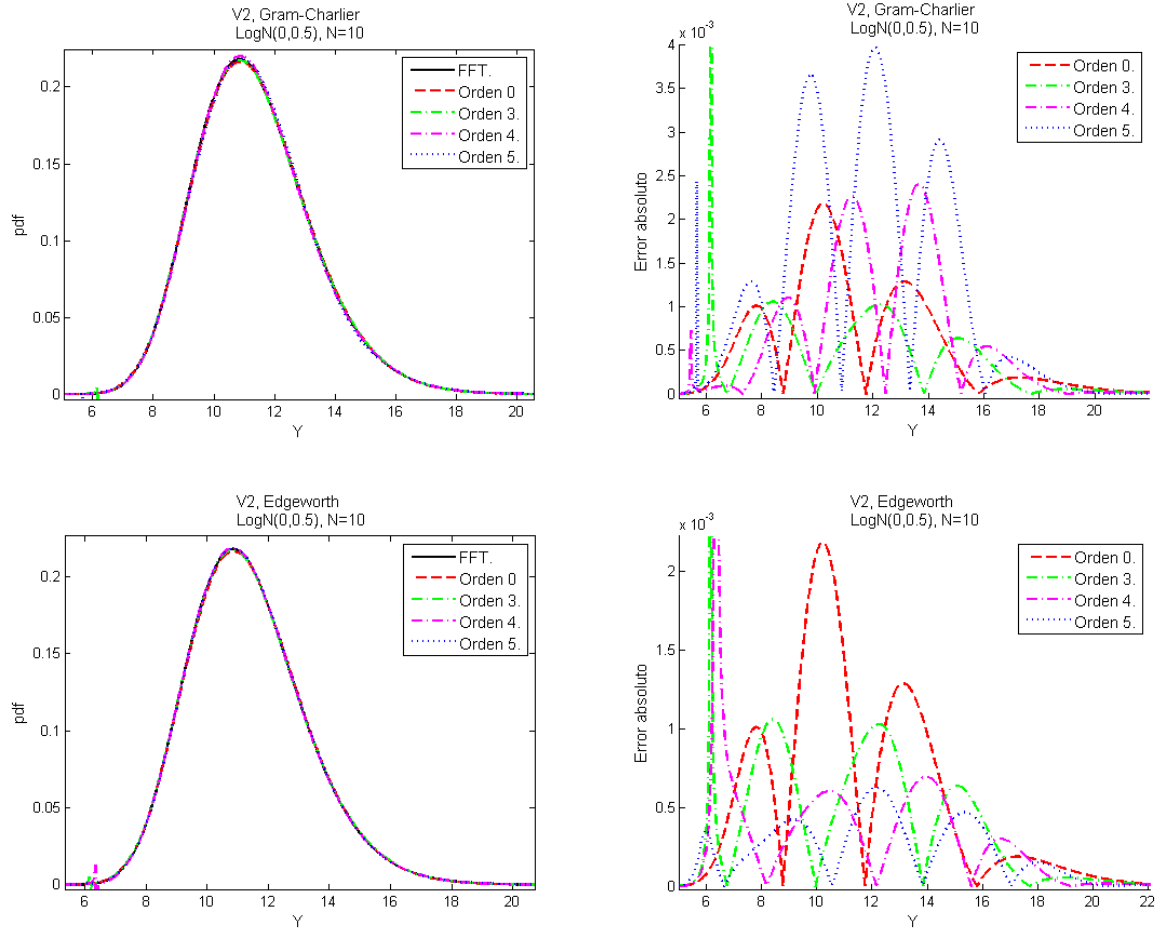
$$f_Y(y) = g_Z(y) \frac{\left(1 + \left(\frac{k_3^{[Y]}}{3!\sigma^{[Y]3}} H_3(\hat{y})\right) + \left(\frac{k_4^{[Y]}}{4!\sigma^{[Y]4}} H_4(\hat{y}) + 10 \frac{k_3^{[Y]2}}{6!\sigma^{[Y]6}} H_6(\hat{y})\right) + \dots\right)}{\left(1 + \left(\frac{k_3^{[Z]}}{3!\sigma^{[Y]3}} H_3(\hat{y})\right) + \left(\frac{k_4^{[Z]}}{4!\sigma^{[Y]4}} H_4(\hat{y}) + 10 \frac{k_3^{[Z]2}}{6!\sigma^{[Y]6}} H_6(\hat{y})\right) + \dots\right)}, \quad (59)$$

donde  $\hat{y} = \left(\frac{y - \mu^{[Y]}}{\sigma^{[Y]}}\right)$  y  $g_Z(y)$  toma su definición de lognormal (50).

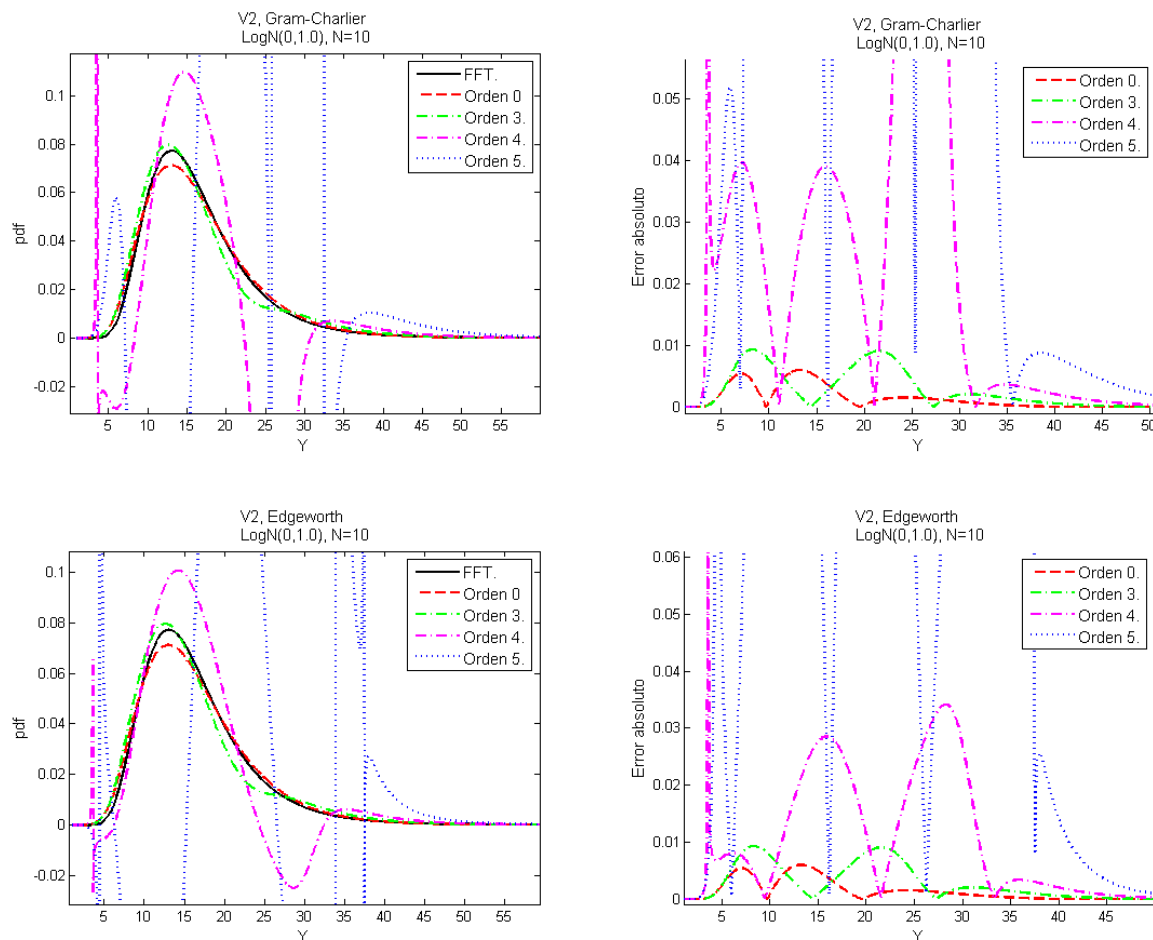
### 5.2.1. Suma de 10 variables



**Figura 9:** Aproximación a la pdf de una suma de 10 variables lognormales con  $\mu = 0$  y  $\sigma = 0.3$ . En la parte superior se emplea Gram-Charlier modificado y en la inferior Edgeworth modificado. De izquierda a derecha se encuentra la pdf y el error absoluto.



**Figura 10:** Aproximación a la pdf de una suma de 10 variables lognormales con  $\mu = 0$  y  $\sigma = 0.5$ . En la parte superior se emplea Gram-Charlier modificado y en la inferior Edgeworth modificado. De izquierda a derecha se encuentra la pdf y el error absoluto.



**Figura 11:** Aproximación a la pdf de una suma de 10 variables lognormales con  $\mu = 0$  y  $\sigma = 1$ . En la parte superior se emplea Gram-Charlier modificado y en la inferior Edgeworth modificado. De izquierda a derecha se encuentra la pdf y el error absoluto.

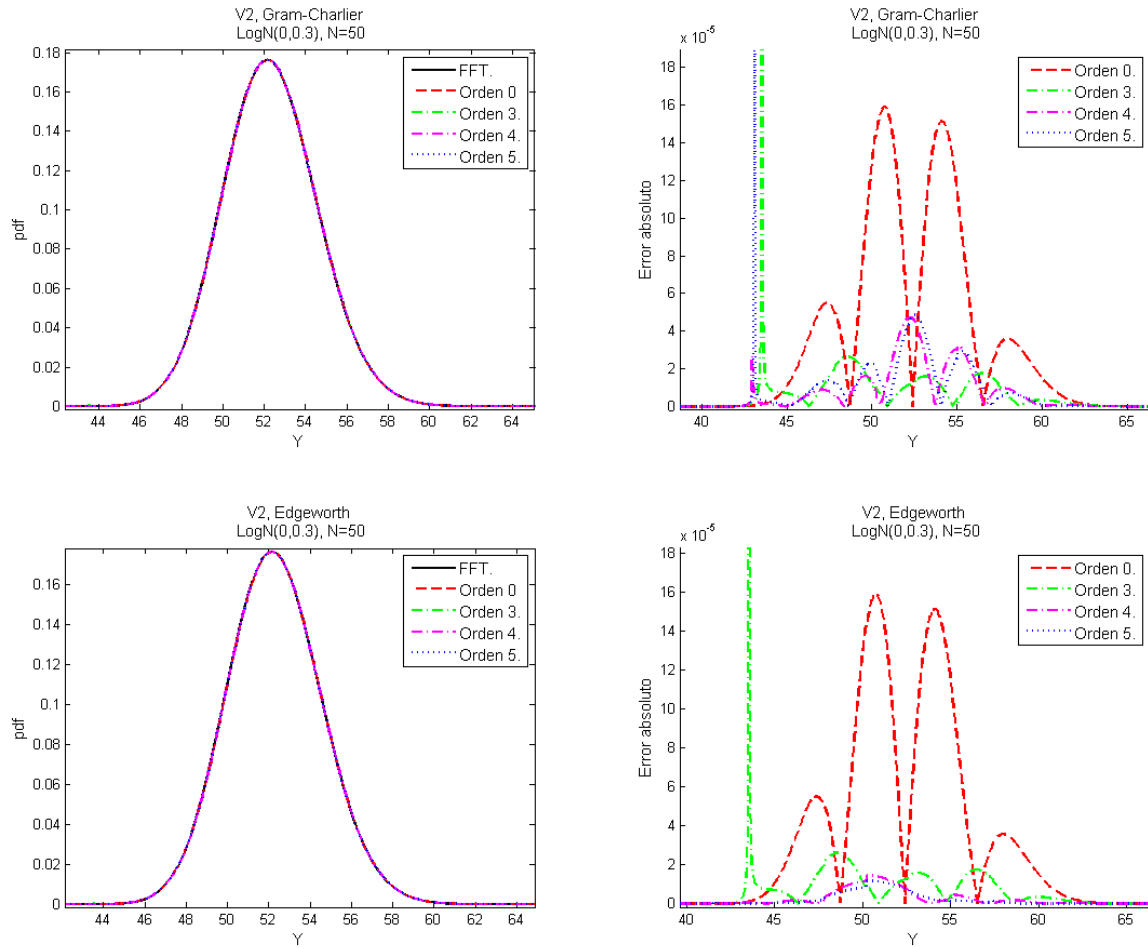
En primer lugar, observemos en las figuras (9), (10) y (11) que para algunas truncaciones justo antes de comenzar el cuerpo de la distribución se producen picos, éstos podemos verlos más claramente en las gráficas de los errores.

Estas anomalías se deben a que las series empleadas (58) y (59) constan de un denominador. Al inicio del cuerpo este denominador toma valores muy cercanos a 0 produciendo así estos indeseados picos.

Al comparar las gráficas obtenidas con Gram-Charlier y con Edgeworth modificados se sigue corroborando que esta última mejora las series de orden 4 y 5. Con la modificación de Edgeworth para  $\sigma < 0.6$  aproximadamente el menor error se tiene con orden 5, sin embargo al aumentar las varianzas se convierte en la peor aproximación, perdiendo incluso la similitud de la forma.

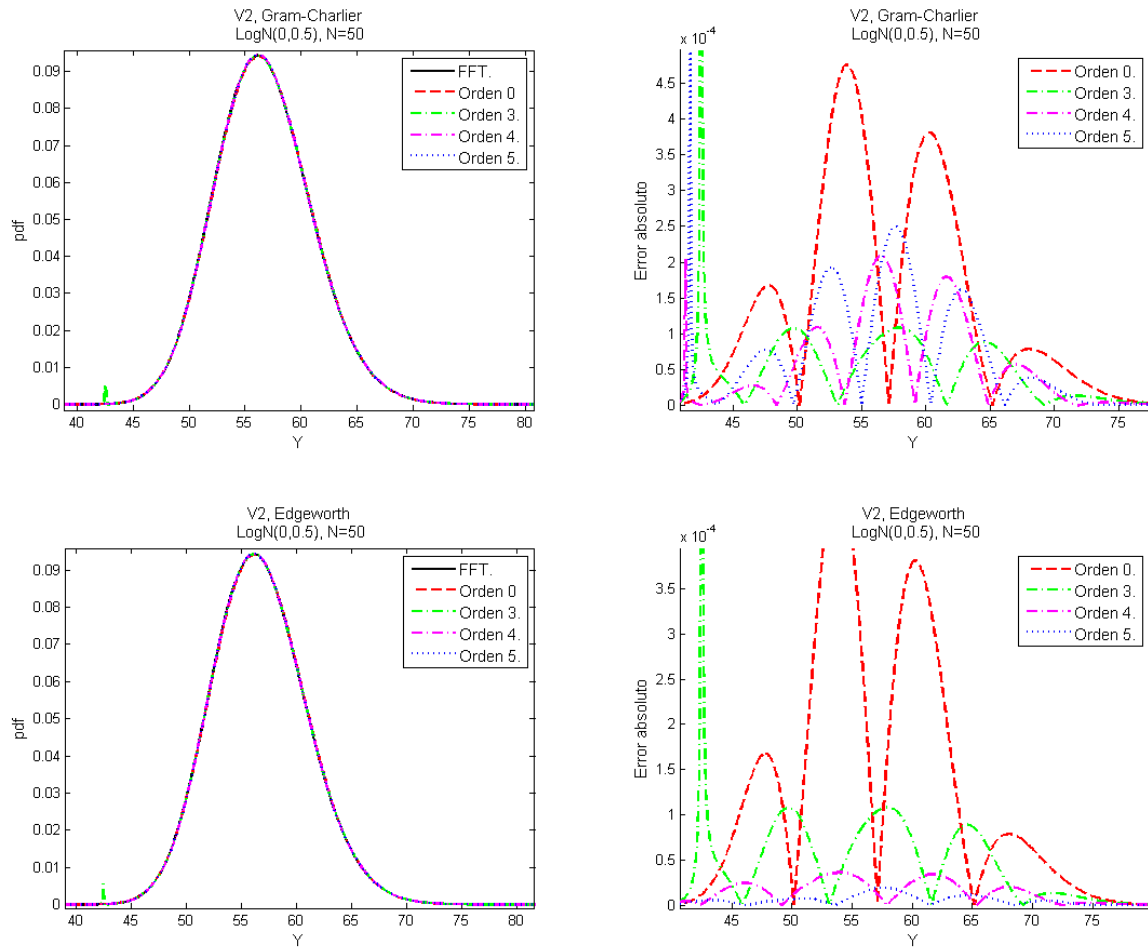
Para  $\sigma = 1$  las soluciones que más se ajustan a la pdf son las de orden 0 y 3. Lo que nos indica que con el método de Fenton-Wilkinson se consiguen buenos resultados, o por lo menos mejores a las series para estos parámetros particulares.

### 5.2.2. Suma de 50 variables

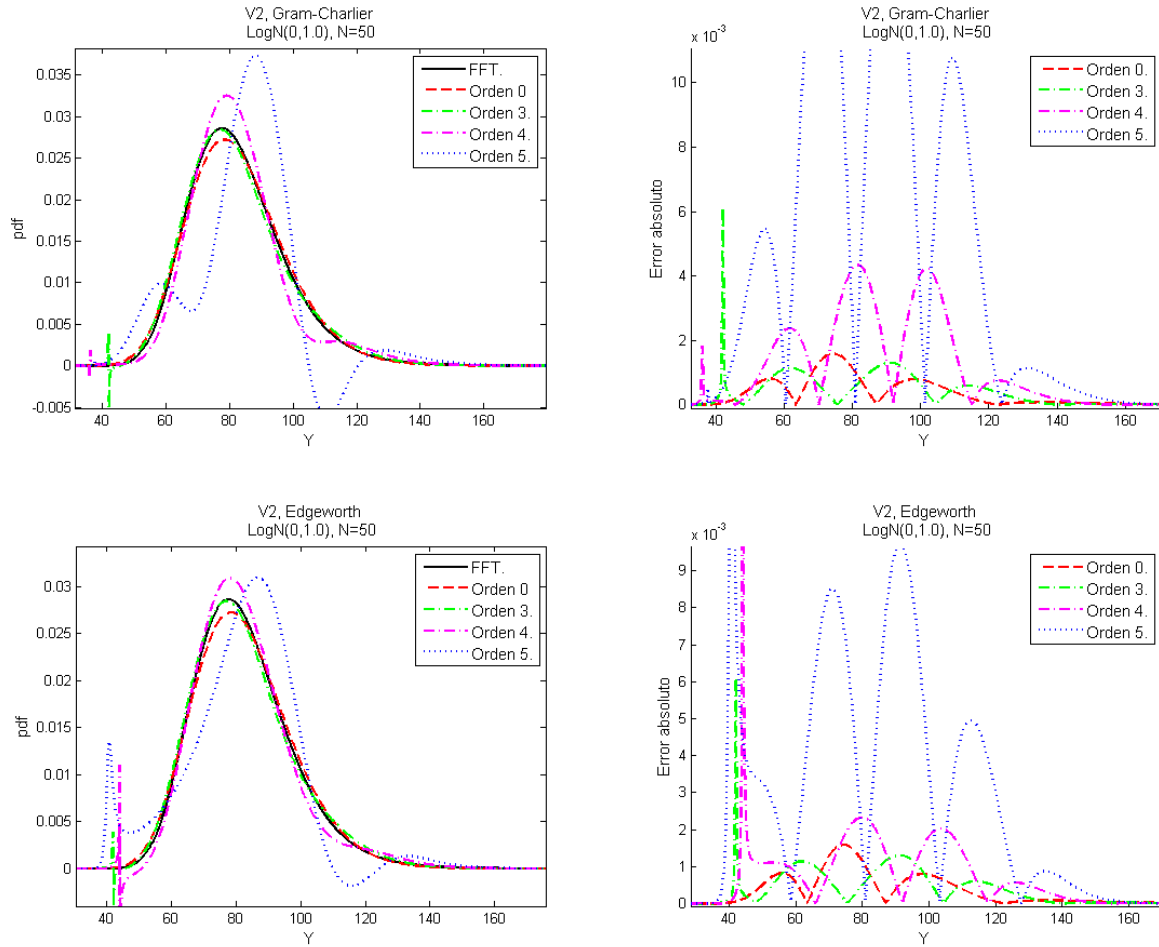


**Figura 12:** Aproximación a la pdf de una suma de 50 variables lognormales con  $\mu = 0$  y  $\sigma = 0.3$ . En la parte superior se emplea Gram-Charlier modificado y en la inferior Edgeworth modificado. De izquierda a derecha se encuentra la pdf y el error absoluto.





**Figura 13:** Aproximación a la pdf de una suma de 50 variables lognormales con  $\mu = 0$  y  $\sigma = 0.5$ . En la parte superior se emplea Gram-Charlier modificado y en la inferior Edgeworth modificado. De izquierda a derecha se encuentra la pdf y el error absoluto.



**Figura 14:** Aproximación a la pdf de una suma de 50 variables lognormales con  $\mu = 0$  y  $\sigma = 1$ . En la parte superior se emplea Gram-Charlier modificado y en la inferior Edgeworth modificado. De izquierda a derecha se encuentra la pdf y el error absoluto.

Nuevamente y esta vez con mayor intensidad, volvemos a encontrar presentes los picos al inicio del cuerpo de la distribución. Eludiendo este suceso, el gran cambio al aumentar el número de variables de la suma es que las series se ajustan con mayor precisión a la solución exacta. Basta con observar las figuras (11) y (14) con  $\sigma = 1$  para apreciarlo.

Comparando ahora con la versión 1, en esta modificación de las series se produce una verdadera mejoría obviando las anomalías ya mencionadas. Dicha mejoría la podemos comprobar con la magnitud del eje de ordenadas del error absoluto y los errores de las mejores truncaciones para cada caso.

### 5.3. Versión 3

Por último en esta versión emplearemos las series generalizadas de Gram-Charlier (22) y Edgeworth (35). Como función de referencia utilizaremos la función lognormal correspondiente a la variable  $Z$ . Tomando como  $Z$  la misma aproximación que en la versión 2,

$$Z \sim \text{LogN}(\mu', \sigma'),$$

y pdf

$$g_z(z) = \frac{1}{z\sigma'\sqrt{2\pi}} e^{-\frac{(\ln(z)-\mu')^2}{2\sigma'^2}}. \quad (60)$$

Versión 3 Gram-Charlier:

$$f_Y(y) = g_z(y) - \frac{(k_3^{[Y]} - k_3^{[Z]})}{3!} D_y^3 g_z(y) + \frac{(k_4^{[Y]} - k_4^{[Z]})}{4!} D_y^4 g_z(y) + \dots \quad (61)$$

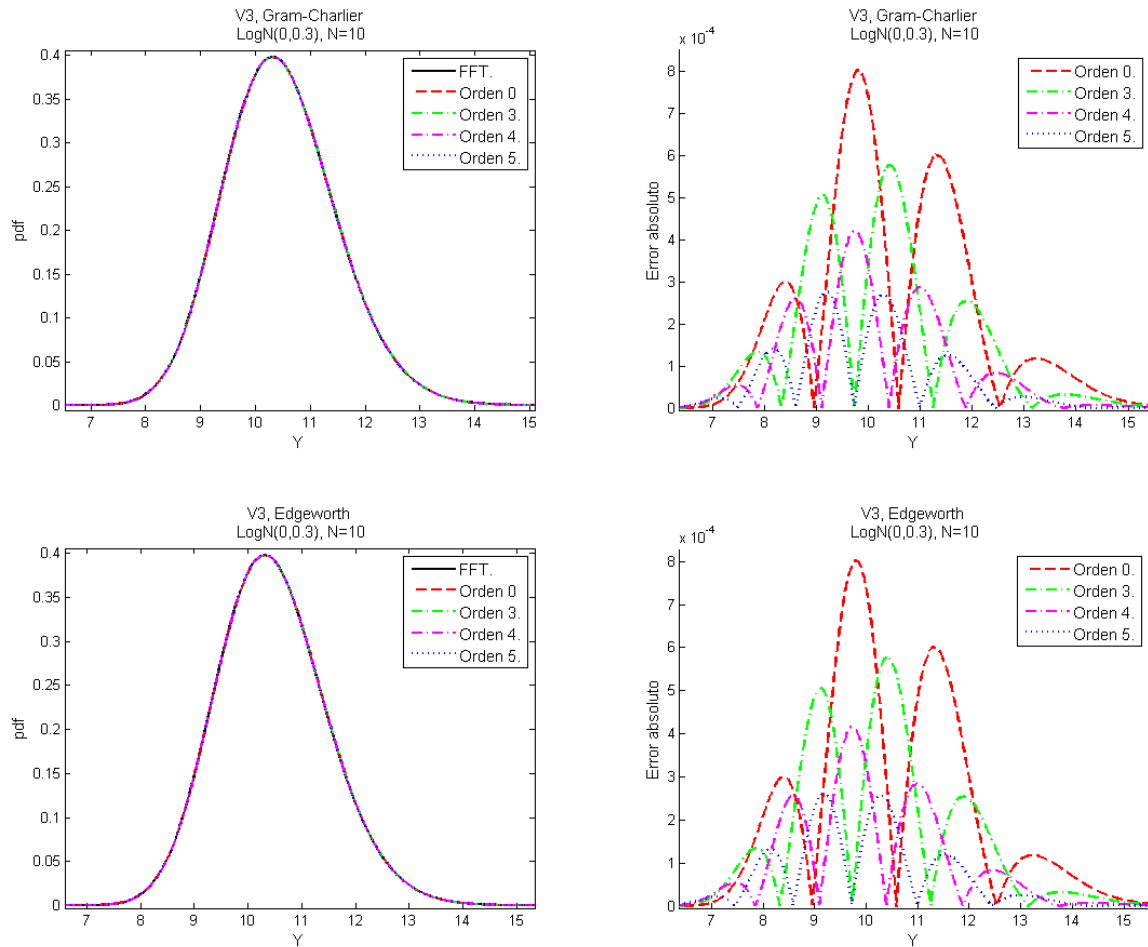
Versión 3 Edgeworth:

$$\begin{aligned} f_Y(y) = & g_z(y) + \\ & + \left( -\frac{k_3^{[Y]} - k_3^{[Z]}}{3!} D_y^3 g_z(y) \right) + \\ & + \left( \frac{k_4^{[Y]} - k_4^{[Z]}}{4!} D_y^4 g_z(y) + 10 \frac{(k_3^{[Y]} - k_3^{[Z]})^2}{6!} D_y^6 g_z(y) \right) + \\ & + \dots \end{aligned} \quad (62)$$

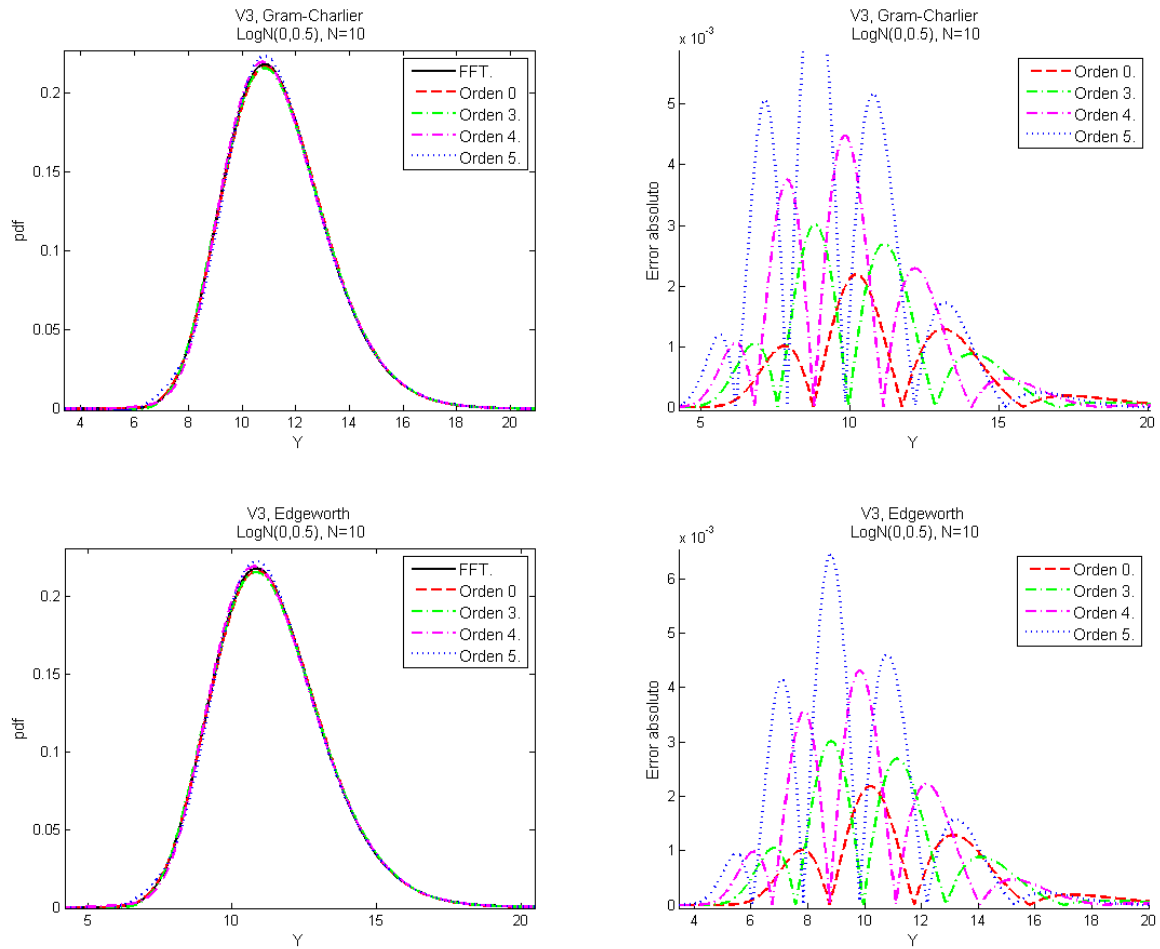
Notemos que tomando únicamente orden 0 se tiene la función  $g_z(y)$  y al elegir la distribución de  $Z$  idéntica a la versión 2 se tendrá que las series modificadas coinciden en orden 0 a las generalizadas.

#### 5.3.1. Suma de 10 variables

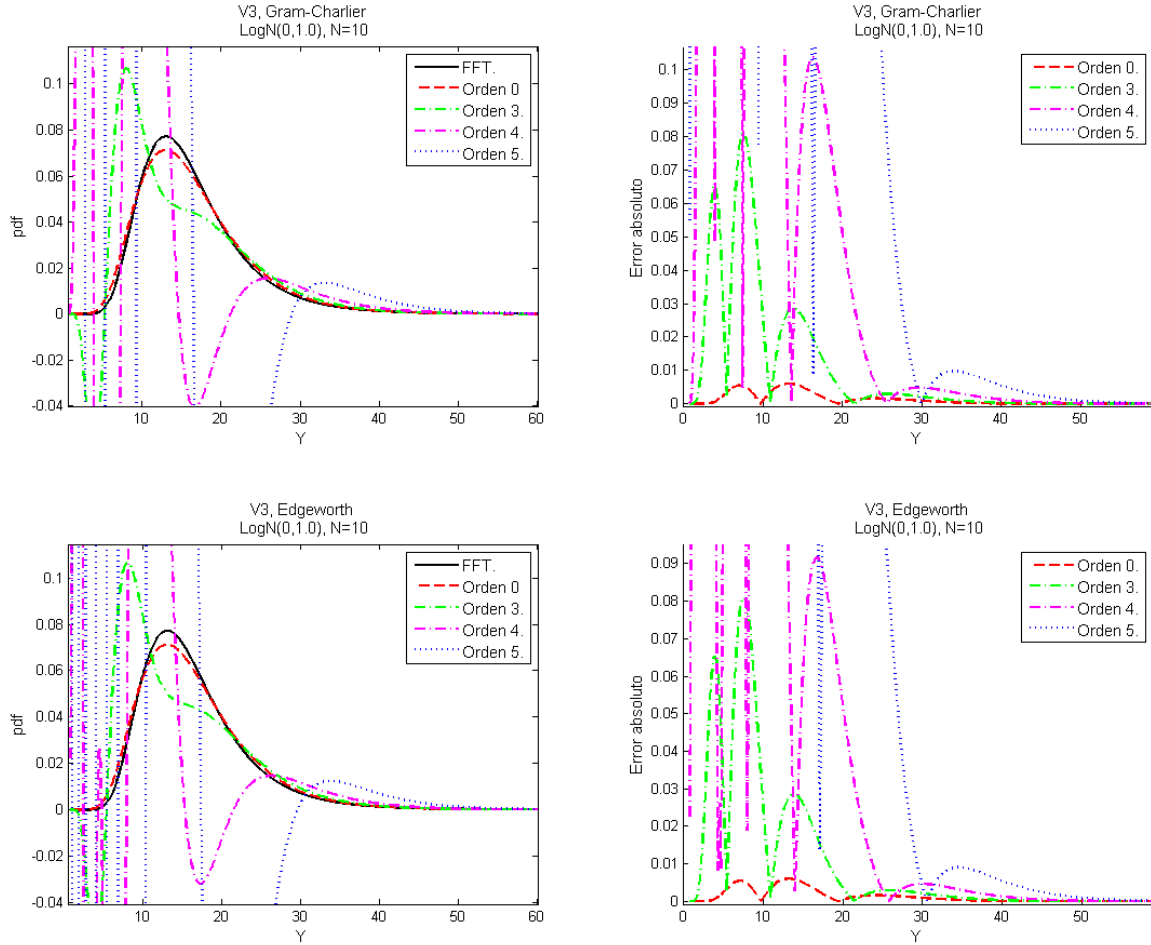
Se presentan a continuación las gráficas correspondientes a la tercer versión para aproximar sumas de diez lognormales con parámetro  $\mu = 0$  fijo y  $\sigma$  tomando los valores 0.3, 0.5 y 1 sucesivamente.



**Figura 15:** Aproximación a la pdf de una suma de 10 variables lognormales con  $\mu = 0$  y  $\sigma = 0.3$ . En la parte superior se emplea Gram-Charlier generalizado y en la inferior Edgeworth generalizado. De izquierda a derecha se encuentra la pdf y el error absoluto.



**Figura 16:** Aproximación a la pdf de una suma de 10 variables lognormales con  $\mu = 0$  y  $\sigma = 0.5$ . En la parte superior se emplea Gram-Charlier generalizado y en la inferior Edgeworth generalizado. De izquierda a derecha se encuentra la pdf y el error absoluto.

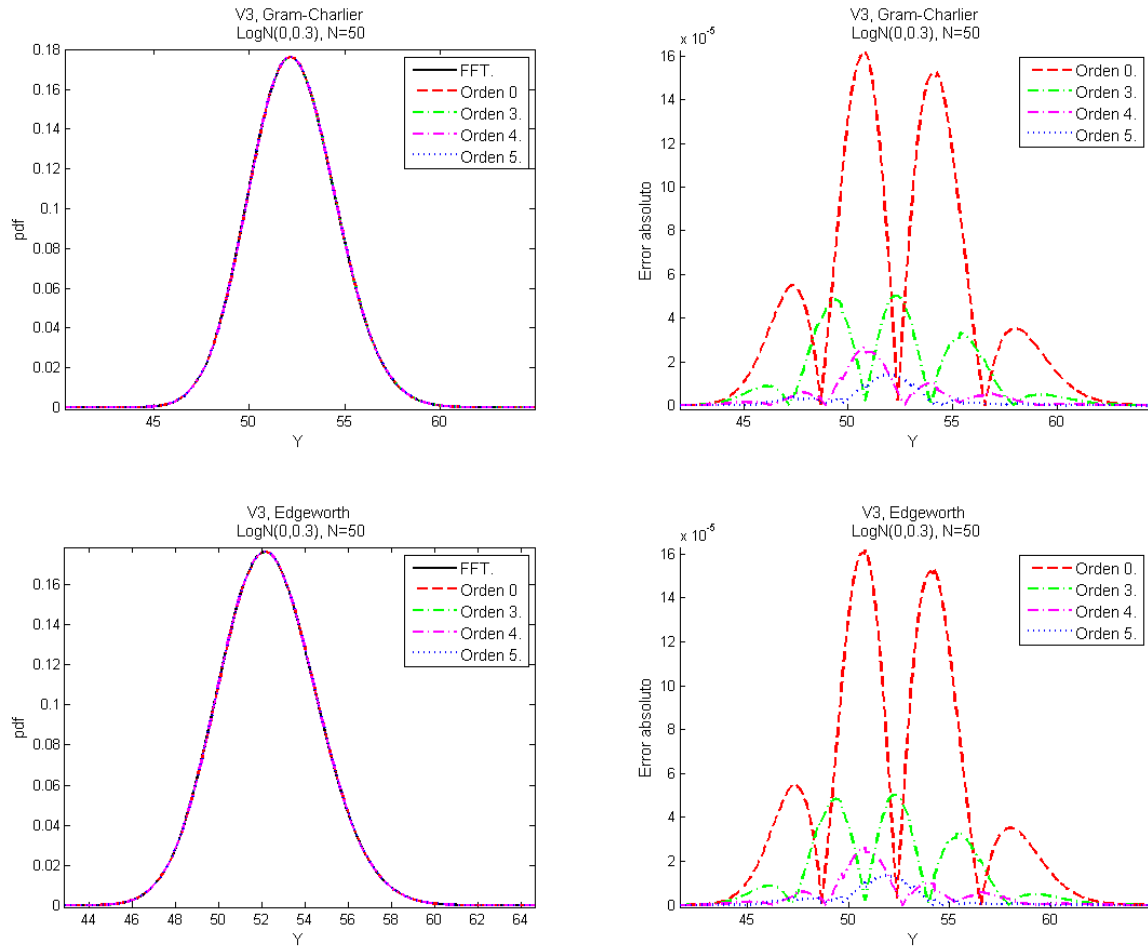


**Figura 17:** Aproximación a la pdf de una suma de 10 variables lognormales con  $\mu = 0$  y  $\sigma = 1$ . En la parte superior se emplea Gram-Charlier generalizado y en la inferior Edgeworth generalizado. De izquierda a derecha se encuentra la pdf y el error absoluto.

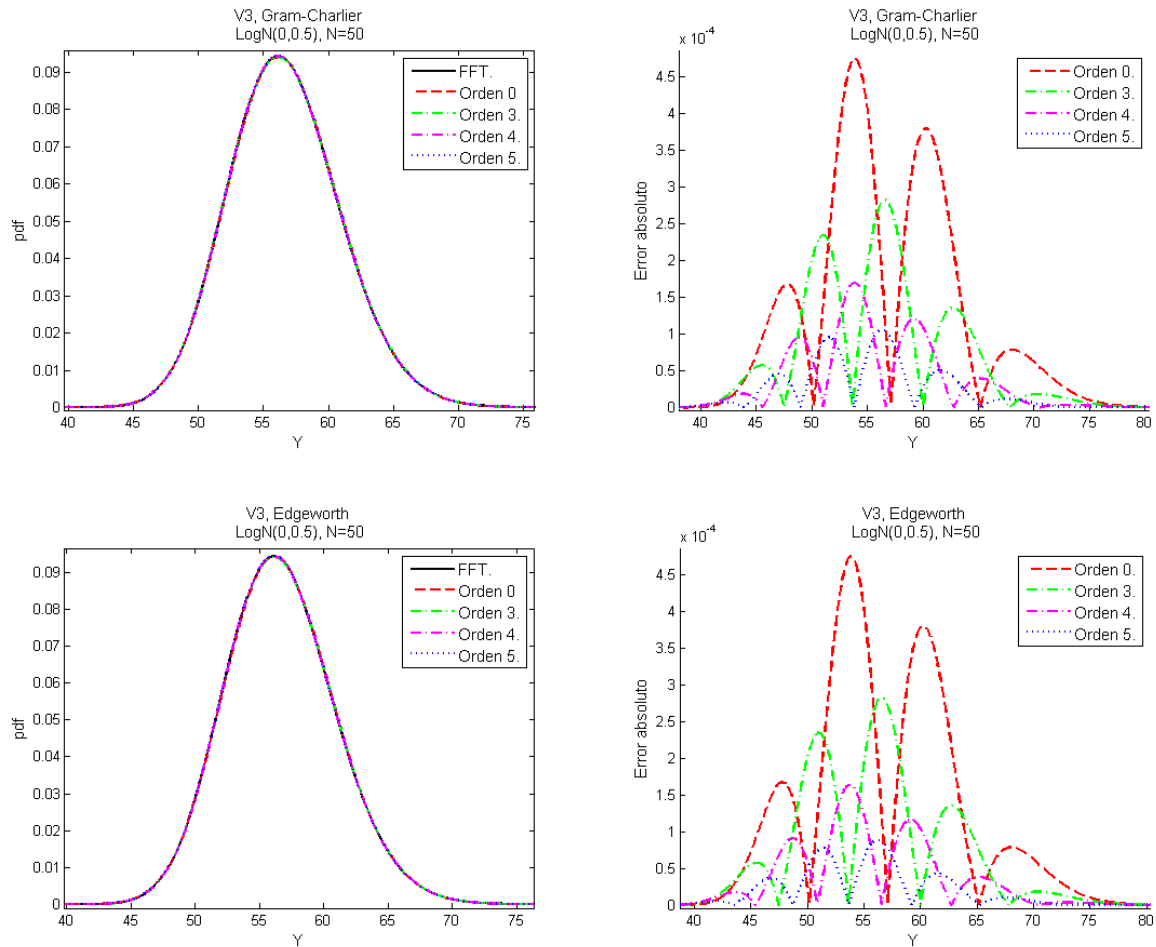
En esta versión y con  $N = 10$  nos encontramos con que las series de Gram-Charlier y Edgeworth generalizadas son muy semejantes, con cambios casi inapreciables.

Apreciamos también que hasta  $\sigma = 0.4$  aproximadamente, la mejor truncación es la de orden 5, si embargo al tomar valores superiores comienza a empeorar drásticamente.

### 5.3.2. Suma de 50 variables

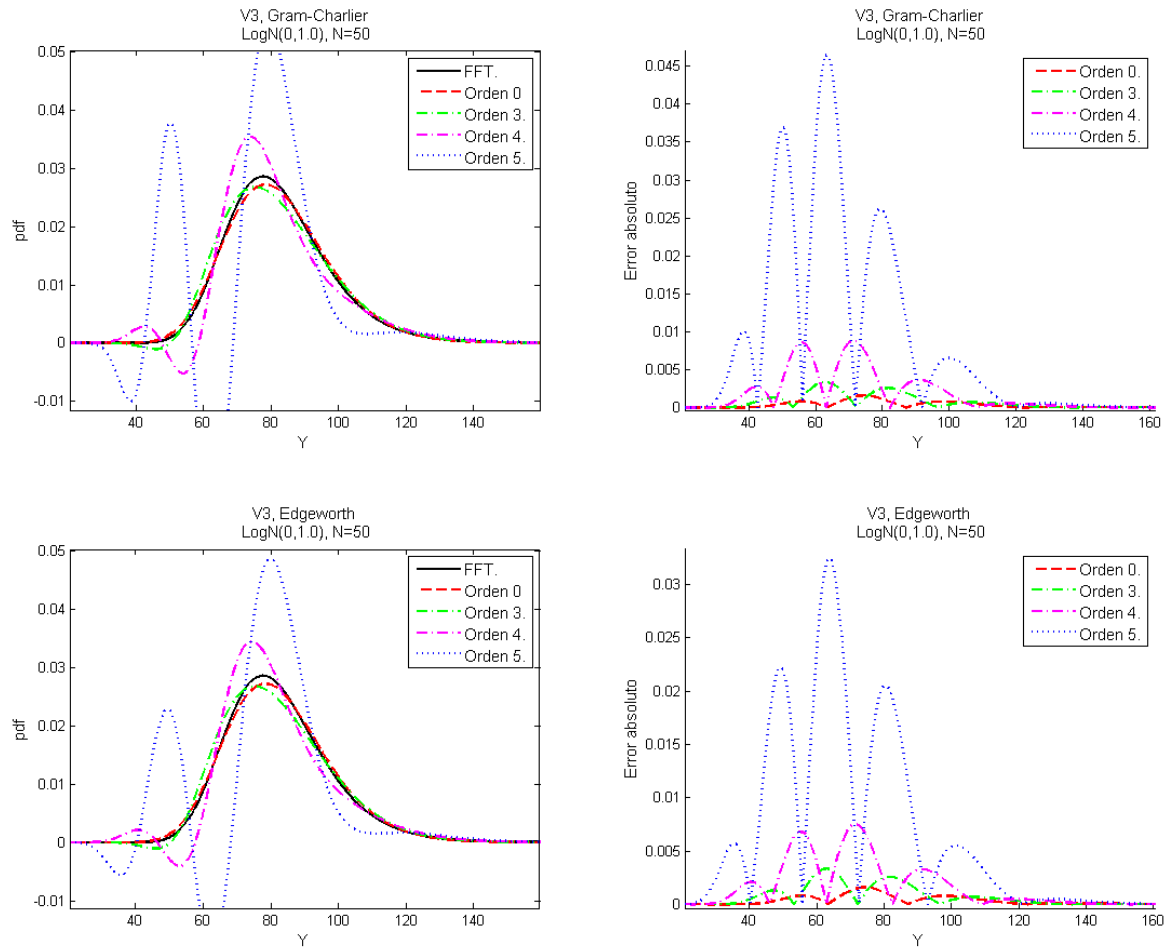


**Figura 18:** Aproximación a la pdf de una suma de 50 variables lognormales con  $\mu = 0$  y  $\sigma = 0.3$ . En la parte superior se emplea Gram-Charlier generalizado y en la inferior Edgeworth generalizado. De izquierda a derecha se encuentra la pdf y el error absoluto.



**Figura 19:** Aproximación a la pdf de una suma de 50 variables lognormales con  $\mu = 0$  y  $\sigma = 0.5$ . En la parte superior se emplea Gram-Charlier generalizado y en la inferior Edgeworth generalizado. De izquierda a derecha se encuentra la pdf y el error absoluto.





**Figura 20:** Aproximación a la pdf de una suma de 50 variables lognormales con  $\mu = 0$  y  $\sigma = 1$ . En la parte superior se emplea Gram-Charlier generalizado y en la inferior Edgeworth generalizado. De izquierda a derecha se encuentra la pdf y el error absoluto.

Al tomar  $N = 50$  apreciamos que las series se ajustan mejor a la pdf que en el caso  $N = 10$  y las series de orden 5 mantienen su semejanza a la solución para mayores valores de  $\sigma$ .

Comparando la versión 3 con la 2, notamos que de la segunda resultan curvas ligeramente mejores. Este hecho se puede apreciar teniendo en cuenta la truncación de orden 0 que coincide en ambos casos, ésta comienza a ser la mejor solución de las series generalizadas antes que en las series modificadas.

Por ejemplo, para  $N = 10$  y  $\sigma = 0.5$  en figura (16) correspondiente a la tercera versión la truncación de orden 0 es mejor, mientras que en la figura (10) de la segunda versión se puede observar que no es la de menor error.



## 6. Conclusiones

### 6.1. Conclusiones y contribuciones

El principal objetivo que tuvo el presente trabajo fue encontrar una aproximación a las distribuciones de sumas de variables aleatorias positivas con colas pesadas. Motivado por sus importantes aplicaciones de interés práctico, donde cabe destacar su uso en las comunicaciones móviles.

Para este fin, se llevó a cabo un estudio sobre aproximaciones obtenidas mediante la truncación de las expansiones de Gram-Charlier y las series asintóticas de Edgeworth. Estas series se han utilizado para aproximar la pdf de sumas de variables aleatorias cuando el número de sumandos no es suficientemente grande como para emplear el teorema del límite central con resultados satisfactorios.

En este estudio se desarrollaron tres aproximaciones distintas basadas en estas series, las cuales se han denominado clásica, generalizada y modificada. Con el fin de determinar la precisión de las distintas aproximaciones se realizó una fase de experimentación centrada en sumas de variables lognormales.

En primer lugar, se presentaron las series clásicas relacionadas con el teorema central de límite al tener como función de referencia la gaussiana. Con ellas se obtuvieron los peores resultados comparando con el resto de modalidades, aunque se observó que la serie de Edgeworth mostraba una mejoría frente a Gram-Charlier.

La generalización de la serie de Gram-Charlier empleada en [10] donde se muestran los buenos resultados obtenidos para aproximar una cdf motivó a la realización adicional del desarrollo para la correspondiente serie de Edgeworth generalizada.

En este caso, sin embargo, Edgeworth no presentó una mejora tan significativa frente a Gram-Charlier. Aunque ambas series resultaron proporcionar una aproximación más precisa a la solución exacta que las clásicas.

Con el objetivo de mejorar la precisión de las aproximaciones, se modificó la expresión de las series clásicas para poder utilizar una función de referencia diferente a la gaussiana, que proporcione una mejor aproximación a orden 0 de la pdf de la suma de lognormales. Estas series modificadas aportan un valor diferencial, obteniendo los mejores resultados en las pruebas prácticas.

En general, para todas las expansiones se concluye que al aumentar el número de variables éstas se ajustan mejor a la solución. Por otra parte, en el caso particular de variables lognormales, con  $\mu = 0$  se aprecia como al elevar el valor de  $\sigma$  las aproximaciones empeoran. Se establece entonces una relación entre  $N$  (número de sumandos) y  $\sigma$  que determina la calidad de las aproximaciones.

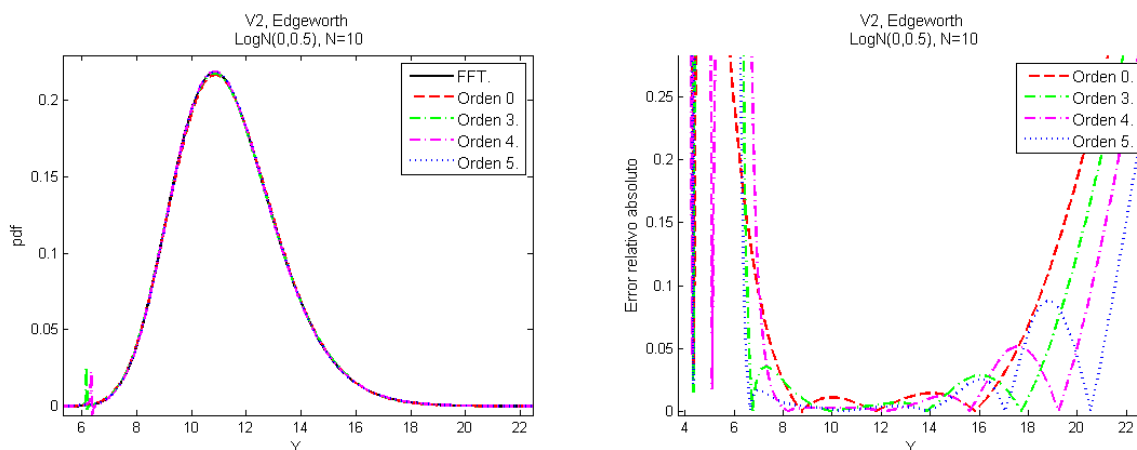
Por otro lado, para la fase de experimentación se han desarrollado programas y una librería en Matlab. Los cuales proporcionan para las distintas series las aproximaciones hasta orden 5 con cualquier número de variables y parámetros de una distribución lognormal. Esta flexibilidad y la cantidad de información que se ofrece permite realizar un estudio detallado teniendo en cuenta todas las variables que influyen en los resultados.

En conclusión, se ha encontrado una modificación de las series originales de Gram-Charlier y Edgeworth que actúa como aproximador de sumas de variables aleatorias, siendo la correspondiente a Edgeworth la que presenta mayor calidad. En el caso particular de log-normales, cuanto más pequeño se toma el valor del parámetro  $\sigma$  los resultados obtenidos tienen menor error, produciendo soluciones satisfactorias de error casi inapreciable.

## 6.2. Trabajo futuro

Gran importancia de las distribuciones que han sido aproximadas recae sobre la propiedad de sus colas pesadas. Por ello se plantea como continuación de este trabajo crear una con aproximación otros métodos distintos a los estudiados para obtener mayor precisión sobre la cola de las distribuciones. De esta manera se realizaría una aproximación por partes, en el cuerpo se emplearía la serie de Edgeworth modificada y en la cola la nueva expansión.

En las gráficas mostradas en la sección de experimentación no se aprecia el error cometido sobre la cola de la distribución de sumas lognormales ya que los valores tomados por la pdf en esta parte son mucho más pequeños que los tomados en el cuerpo. Sin embargo, al tomar el error relativo, éste es mayor sobre la cola que sobre el cuerpo. En la figura (21) se puede observar este hecho.



**Figura 21:** Aproximación a la pdf de una suma de 10 variables lognormales con  $\mu = 0$  y  $\sigma = 0.5$ . De izquierda a derecha se encuentra la pdf y el error relativo absoluto.

Para aproximar las colas de las distribuciones de sumas de variables aleatorias positivas con colas pesadas se propone emplear un desarrollo perturbativo. En [7] y [6] se estu-

dian estos desarrollos para la aproximación de la cdf, habría entonces que calcular el correspondiente desarrollo para la pdf.



## A. Cumulantes

En probabilidad, los cumulantes de una variable aleatoria vienen determinados por los momentos correspondientes a la misma. Éstos son empleados en las series de Gram-Charlier y Edgeworth como veremos posteriormente.

Veamos la definición de la función generatriz de momentos, la cual se puede expresar mediante una serie, utilizando el desarrollo de Taylor correspondiente a la función exponencial:

$$M_Y(t) := E(e^{tY}) = E\left(\sum_{m=0}^{\infty} \frac{t^m}{m!} Y^m\right) = 1 + \sum_{m=0}^{\infty} \frac{t^m}{m!} E(Y^m) \quad (63)$$

Observemos que al disponer de la función generadora de momentos derivándola podemos obtener cualquiera de los momentos no centrales, siempre que esta exista en un entorno de  $t = 0$ :

$$\left. \frac{d^n}{dt^n} M_Y(t) \right|_{t=0} = \left. \frac{d^n}{dt^n} \left( 1 + \sum_{m=0}^{\infty} \frac{t^m}{m!} E(Y^m) \right) \right|_{t=0} = E(Y^n) = \mu^n \quad (64)$$

Al igual que la función generadora de momentos es una serie creada en base a sus momentos, existe una respectiva función generatriz de cumulantes y se define como el logaritmo de  $M_Y(t)$ .

$$G_Y(t) := \log(M_Y(t)) := \log(E(e^{tY})) \quad (65)$$

$$G_Y(t) = \sum_{m=1}^{\infty} \frac{t^m}{m!} k_m, \quad (66)$$

donde  $k_m$  denota el cumulante de orden  $m$ . En ocasiones veremos  $k_m^{[Y]}$ , para especificar que es un cumulante correspondiente a la v.a.  $Y$ .

De igual forma que obteníamos el momento de orden  $n$  derivando  $n$  veces  $M_Y(t)$  y evaluando en  $t = 0$ , podemos obtener el cumulante  $n$ -ésimo, sin olvidar que para ello  $g_Y(t)$  debe estar definida en un entorno de  $t = 0$ :

$$\left. \frac{d^n}{dt^n} G_Y(t) \right|_{t=0} = \left. \frac{d^n}{dt^n} \left( \sum_{m=1}^{\infty} \frac{t^m}{m!} k_m \right) \right|_{t=0} = k_n \quad (67)$$

Una propiedad muy útil de los cumulantes al trabajar con sumas de variables aleatorias es que son aditivos. Sea  $Y$  una v.a. definida como

$$Y = \sum_{j=1}^N X_j$$

con  $\{X_j\}_1^N$  v.a. independientes. Entonces, empleando la definición de la función generadora de cumulantes (65),

$$\begin{aligned} G_Y(t) &:= \log\left(E(e^{tY})\right) = \log\left(E\left(\prod_{j=1}^N e^{tX_j}\right)\right) = \log\left(\prod_{j=1}^N E(e^{tX_j})\right) \\ &= \sum_{j=1}^N \log\left(E(e^{tX_j})\right) = \sum_{j=1}^N \log\left(E(e^{tX_j})\right) = \sum_{j=1}^N G_{X_j}(t) \end{aligned} \quad (68)$$

tenemos que  $G_Y(t)$  es igual a la suma de las correspondientes funciones generadoras de cumulantes de  $X_j$ , por lo que al derivar a ambos lados de la igualdad empleando (67) conseguimos el resultado esperado:

$$k_n^{[Y]} = \sum_{j=1}^N k_n^{[X_j]}, \quad \forall n \in \mathbb{R} \quad (69)$$

Si las v.a.i.  $\{X_j\}_1^N$  estuviesen además idénticamente distribuidas:

$$k_n^{[Y]} = N k_n^{[X_1]}, \quad \forall n \in \mathbb{R} \quad (70)$$

Como ya hemos visto, los cumulantes quedan definidos mediante los momentos no centrales, una relación para encontrar los cumulantes en función de momentos y cumulantes anteriores es la siguiente:

$$k_n = \mu_n - \sum_{m=1}^{n-1} \binom{n-1}{m-1} k_m \mu_{n-m} \quad (71)$$

### Demostración

Derivamos la definición (65) respecto de  $t$  y reordenamos la ecuación resultante:

$$\frac{d}{dt}\left(M_Y(t)\right) = M_Y(t) \frac{d}{dt}\left(G_Y(t)\right) \quad (72)$$

Ahora derivamos  $n-1$  veces (72), teniendo en cuenta  $(f(x)g(x))^{(n)} = \sum_{l=0}^n \binom{n}{l} f^{(l)}(x) g^{(n-l)}(x)$ :

$$\frac{d^n}{dt^n}\left(M_Y(t)\right) = \sum_{l=0}^{n-1} \binom{n-1}{l} \frac{d^{n-1-l}}{dt^{n-1-l}}\left(M_Y(t)\right) \frac{d^{l+1}}{dt^{l+1}}\left(G_Y(t)\right) \quad (73)$$



Evaluamos en  $t = 0$  (73), recordando (64) y (67) y realizamos el cambio de variable  $l = m - 1$ :

$$\mu_n = \sum_{l=0}^{n-1} \binom{n-1}{l} \mu_{n-1-l} k_{l+1} = k_n + \sum_{m=1}^{n-1} \binom{n-1}{m-1} \mu_{n-m} k_m$$

Despejando  $k_n$  obtenemos la ecuación buscada:

$$k_n = \mu_n - \sum_{m=1}^{n-1} \binom{n-1}{m-1} \mu_{n-m} k_m$$

■

Con esta fórmula recursiva podemos encontrar fácilmente los cumulantes en función de los momentos anteriores, los primeros cuatro son:

$$\begin{aligned} k_1 &= \mu_1, \\ k_2 &= \mu_2 - k_1 \mu_1 = \mu_2 - \mu_1^2, \\ k_3 &= \mu_3 - k_1 \mu_2 - 2k_2 \mu_1 = \mu_3 - 3\mu_1 \mu_2 + 2\mu_1^3, \\ k_4 &= \mu_4 - k_1 \mu_3 - 3k_2 \mu_2 - 3k_3 \mu_1 = \mu_4 - 4\mu_1 \mu_3 - 3\mu_2^2 + 12\mu_1^2 \mu_2 - 6\mu_1^4. \end{aligned} \tag{74}$$

## B. Polinomios de Hermite

Los polinomios de Hermite son un conjunto de polinomios ortogonales, los utilizados en este trabajo son habitualmente llamados polinomios de Hermite probabilísticos:

$$H_n(x) = \frac{(-1)^n}{\phi(x)} \frac{d^n}{dx^n} \phi(x) \quad (75)$$

donde  $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ , función de densidad de probabilidad de la distribución normal estándar.

Los diez primeros polinomios explícitos son:

$$\begin{aligned} H_0(x) &= 1, \\ H_1(x) &= x, \\ H_2(x) &= x^2 - 1, \\ H_3(x) &= x^3 - 3x, \\ H_4(x) &= x^4 - 6x^2 + 3, \\ H_5(x) &= x^5 - 10x^3 + 15x, \\ H_6(x) &= x^6 - 15x^4 + 45x^2 - 15, \\ H_7(x) &= x^7 - 21x^5 + 105x^3 - 105x, \\ H_8(x) &= x^8 - 28x^6 + 210x^4 - 420x^2 + 105, \\ H_9(x) &= x^9 - 36x^7 + 378x^5 - 1260x^3 + 945x. \end{aligned} \quad (76)$$

Estos polinomios son ortogonales respecto a  $\phi(x)$ , es decir,

$$\int_{-\infty}^{\infty} H_n(x) H_m(x) \phi(x) dx = n! \delta_{n,m} \quad (77)$$

donde  $\delta_{n,m}$  es la Delta de Kronecker,  $\begin{cases} 1, & \text{si } n = m \\ 0, & \text{si } n \neq m \end{cases}$

Cuando trabajamos con la función  $\phi(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$  en vez de con  $\phi(x)$  puede resultar interesante utilizar la siguiente igualdad:

$$(-1)^n \frac{d^n}{dx^n} \phi(x, \mu, \sigma) = \frac{1}{\sigma^n} H_n\left(\frac{x-\mu}{\sigma}\right) \phi(x, \mu, \sigma), \quad (78)$$

### Demostración

$$\frac{(-1)^n}{\phi(x, \mu, \sigma)} \frac{d^n}{dx^n} \phi(x, \mu, \sigma) = \frac{\sigma\sqrt{2\pi}}{\sigma\sqrt{2\pi}} \frac{(-1)^n}{e^{-\frac{(x-\mu)^2}{2\sigma^2}}} \frac{d^n}{dx^n} \left( e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right) \underbrace{\quad}_{z = (x-\mu)/\sigma}$$

cambio de variable  $z = (x - \mu)/\sigma$

$$= \frac{(-1)^n}{e^{-\frac{z^2}{2}} \sigma^n} \frac{d^n}{dx^n} \left( e^{-\frac{z^2}{2}} \right) = \frac{1}{\sigma^n} H_n(z) = \frac{1}{\sigma^n} H_n\left(\frac{x-\mu}{\sigma}\right)$$

■

## C. Delta de Dirac

La Delta de Dirac es una función que se obtiene mediante las transformadas de Fourier directa e inversa. Veamos en primer lugar ambas transformadas y su relación con la Delta de Dirac.

Sea  $f$  una función  $L_2$  la transformada de Fourier directa se define como:

$$\hat{f}(w) = \int_{-\infty}^{\infty} f(t)e^{-iwt}dt \quad (79)$$

Y la transformada de Fourier inversa es:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(w)e^{iwt}dw \quad (80)$$

Sustituyendo en la ecuación anterior  $\hat{f}(w)$  por su definición (79),

$$\begin{aligned} f(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(w)e^{iwt}dw = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} f(t')e^{-iwt'}dt' \right) e^{iwt}dw \\ &= \int_{-\infty}^{\infty} \left( \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{iw(t-t')}dw \right) f(t')dt' = \int_{-\infty}^{\infty} \delta(t-t')f(t')dt' \end{aligned} \quad (81)$$

La función Delta de Dirac queda definida como:

$$\delta(t-t') = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{iw(t-t')}dw \quad (82)$$

$$\delta(t-t') = \begin{cases} \infty, & \text{si } t = t' \\ 0, & \text{si } t \neq t' \end{cases} \quad (83)$$

Dado que  $\delta$  es una función par respecto a cero, la ecuación (81) la veremos escrita habitualmente como:

$$f(t) = \int_{-\infty}^{\infty} \delta(t'-t)f(t')dt' \quad (84)$$

## Referencias

- [1] S. Anatolyev and N. Gospodinov. *Methods for Estimation and Inference in Modern Econometrics*. Taylor & Francis, 2011.
- [2] Krishna Balasundaram Athreya and Soumen N. Lahiri. *Measure theory and probability theory*. Springer Texts in Statistics. Springer, New York, 2006.
- [3] Prakash Balachandran. The Edgeworth Expansion and Convergence in the Central Limit Theorem. 2009.
- [4] Domingo Israel Cruz Báez. Teoría de operadores y cálculo estocástico aplicado a las finanzas. Master's thesis, Universidad de La Laguna, 2008.
- [5] Michiel de Kock. Gram-Charlier and Edgeworth expansions for nongaussian correlations in femtoscopy. *Winter School on Heavy Ion Physics*, 2009.
- [6] Lorenzo Hernández, Jorge Tejero, Alberto Suárez, and Santiago Carrillo-Menéndez. Closed-form approximations for operational var. *Statistics and Computing*, 2013.
- [7] Lorenzo Hernández, Jorge Tejero, Alberto Suárez, and Santiago Carrillo-Menéndez. Percentiles of sums of heavy-tailed random variables: Beyond the single-loss approximation. *Statistics and Computing*, 2013.
- [8] Ray Popovic and David Goldsman. Easy Gram-Charlier valuations of options. *The Journal of Derivatives*, 2012.
- [9] JCS Santos Filho, P Cardieri, and MD Yacoub. Simple accurate lognormal approximation to lognormal sums. *Electronics Letters*, 41(18):1016–1017, 2005.
- [10] D. C. Schleher. Generalized Gram-Charlier series with application to the sum of lognormal variates. *IEEE Transactions on Information Theory*, 23(2):275–280, 1977.
- [11] Gordon L. Stüber. *Principles of Mobile Communication (2Nd Ed.)*. Kluwer Academic Publishers, Norwell, MA, USA, 2001.
- [12] Max Welling. Robust series expansions for probability density estimation. Technical report, California Institute of Technology, Computational Vision Lab, 1999.